

# **Proceedings from the 1st Conference of Pronunciation in Second Language Learning and Teaching**

---

Edited by

**John Levis**

**Kimberly LeVelle**

**2009**

## ACKNOWLEDGEMENTS

The papers in this volume were presented at the first Pronunciation in Second Language Learning and Teaching conference, held at Iowa State University on September 17 – 19 of 2009.

We thank the invited speakers for their enthusiastic participation: Tracey Derwing, Wayne Dickerson, and Robert Godwin-Jones. We also thank the Intensive English and Orientation Program, whose financial support helped to make these proceedings a reality, as well as the Department of English at Iowa State University for their financial support.

Copyright for abstracts and papers written for the First Annual Conference on Pronunciation in Second Language Learning and Teaching (PSLLT) is retained by the individual author/s, who should be contacted for permission by those wishing to use the materials for purposes other than those in accordance with fair use provisions of U.S. Copyright Law.

2010  
TESL/Applied Linguistics  
Iowa State University

## TABLE OF CONTENTS

### Introduction

Rebuilding a Professional Space for Pronunciation.....	1
<i>John Levis and Kimberly LeVelle</i>	

### Part I. Plenary Addresses

Walking the Walk: Integrating the Story of English Phonology .....	9
<i>Wayne Dickerson</i>	

Utopian Goals for Pronunciation Teaching .....	24
<i>Tracey Derwing</i>	

### Part II. Intelligibility, Comprehensibility, and Accentedness

The Effects of Phonetics Training on the Intelligibility and Comprehensibility of Native Spanish Speech by Second Language Learners .....	38
<i>Jennifer Rasmussen and Mary Zampini</i>	

Factors in Raters' Perceptions of Comprehensibility and Accentedness.....	53
<i>Heesung Grace Jun and Jinrong Li</i>	

### Part III. Acquisition and Attitudes

The Effects of Self-Monitoring Strategy Use on the Pronunciation of Learners of English.....	67
<i>Sue Ingels</i>	

A Longitudinal Investigation of Vowel Acquisition .....	90
<i>Fran Gulinello</i>	

ESL Learners' Attitudes toward Pronunciation Instruction and Varieties of English.....	105
<i>Okim Kang</i>	

### Part IV. Descriptive Research

The Pronunciation of <-ED> in Coda Clusters in Somali-Accented English.....	119
<i>Ettien Koffi</i>	

Authentic Speech and Teaching Sentence Focus.....	135
<i>Greta Muller Levis and John Levis</i>	

Levis, J. & LeVelle, K. (2010). Rebuilding a professional space for pronunciation. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1<sup>st</sup> Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 1-9), Ames, IA: Iowa State University.

---

## **Rebuilding a Professional Space for Pronunciation**

**John Levis**

**Kimberly LeVelle**

Iowa State University

The [Pronunciation in Second Language Learning and Teaching \(PSLLT\) conference](#) is the only conference of its kind in North America. This conference came from an idea that germinated during John's sabbatical in Vancouver, British Columbia during the first half of 2008. He took part in one day of presentations and discussions whose goal was to map out research goals and directions, with the end goal of helping to influence the long-term research agenda for pronunciation research and teaching. This led to the idea of a conference devoted to pronunciation teaching and research, and resulted in the planning of this first conference.

A further framing of the reason for the conference came when John was invited to take part in a colloquium at the Canadian Association of Applied Linguistics conference in Ottawa during 2009. The colloquium, *Accentuating the positive: Directions in pronunciation research*, examined many of the same issues as the Vancouver discussion, albeit in greater depth. His part of the colloquium defined the future of pronunciation teaching and research as a marketing issue, an issue of rebranding pronunciation for today's "market". In business, rebranding is a way of reviving demand for an established product that has lost much of its sales appeal in a changing market. Rebranding can occur by developing new uses for a product, showing how it remains relevant to a market that has overlooked its virtues, and building a new brand identity. There are features in language teaching today that suggest that such rebranding could be successful for pronunciation, but there are other elements that indicate that rebranding will not be enough without rebuilding the infrastructure needed to support pronunciation's role in the curriculum.

A positive sign for pronunciation's future is that teachers have not lost interest in teaching pronunciation, nor have students lost interest in learning it. Sessions about teaching pronunciation at professional conferences are routinely jammed. Professional workshops at TESOL, the annual international Teachers of English to Speakers of Other Languages convention, are consistently among the most attended professional development workshops, despite the significant extra cost to participants. Students also recognize pronunciation's value. When we offer pronunciation tutoring opportunities as part of a graduate-level class on the teaching of pronunciation, we can usually accommodate 15 or so students. It is not unusual to have between 100 and 200 students ask to take part. Clearly, learners of English have not heard that pronunciation is unimportant.

---

Pronunciation, as all language teachers with more than a passing knowledge of methodology know, was once central to language teaching. Practical phonetics and phonology were important enough to be part of any well-considered training course. Loss of value to the brand of pronunciation began most obviously with Critical Period research in the 1960s (see a review in Scovel, 2000). This research suggested that native accents were not only unrealistic, but perhaps unachievable for adult learners of a foreign language. Following this, the advent of the Communicative approach to language teaching in the 1970s began to institutionalize the loss of market share for pronunciation because CLT's emphasis on spoken language and communicative effectiveness did not include work on the details of pronunciation accuracy. As a result, in many places the baby (the need for spoken intelligibility) was thrown out with the bathwater (the goal of native accuracy).

By the time research began to recognize that pronunciation was not an issue of native-like vs. unnecessary (Hinofotis & Bailey, 1981), but rather that pronunciation training was essential to a multitude of intermediate steps that influenced spoken intelligibility, the language teaching world had moved on. The infrastructure for the teaching of pronunciation was in serious disrepair with little interest from many in restoring it. Courses for teachers existed in a minority of TESL training programs (Murphy, 1997) in North America; the same situation was true across the English-speaking world (Gilbert, 2010).

Another part of the professional infrastructure, pronunciation research reported in top refereed research journals, was also suffering from neglect. Even today, this is true. A recent survey of research in 14 top professional journals showed that over a 10-year period, from 1999-2008, pronunciation-oriented articles ranged from a low of less than 1 percent to a high of around seven percent of all articles published in these journals (Deng et al., 2009). The highest percentages (none very high) occurred in journals that had published dedicated special topics issues on pronunciation. Several journals went for five years at a time without a single article relevant to pronunciation, indicating that even those teachers looking for research help in making pedagogical decisions were left with few places to turn.

Another part of the professional infrastructure included sanctioned settings for professional to meet. In the International Association of Teachers of English as a Foreign Language (IATEFL), a special Interest Groups (SIG) has existed for pronunciation since the mid 1980s, but getting an equivalent group started in TESOL was met with roadblocks of all sorts. Pronunciation was not considered an important professional undertaking, and Interest Sections focusing on particular language skills were not considered appropriate to an organization that preferred to think of itself in terms of the context, not the content, of teaching (Gilbert, 2010).

An interesting analogy to pronunciation is the teaching of grammar. Also as a result of the communicative revolution in language teaching, grammar was threatened with marginalization. Krashen (e.g., 1985) argued that grammar would naturally develop with the right approach to language teaching. Fortunately for students and for teachers, the mistake of

---

removing grammar from teacher training programs did not progress as far as it did with pronunciation, and it became quickly evident that language learning was unlikely to lead to the attainment that, without explicit grammar teaching, students needed to achieve. In addition, grammar teaching was rebranded with the help of SLA theories of input processing (Vanpatten, 2004) and noticing (Long, 1990) and the development of new approaches to teaching such as Larsen-Freeman's *Form-Meaning-Use* paradigm (2001) and cognitive approaches to teaching grammar (Fotos, 2001).

Pronunciation's place in research and teaching, meanwhile, has had similar theoretical advances and teaching paradigm shifts. Research into pronunciation has been strongly influenced by extensive research by Murray Munro and Tracey Derwing (e.g., 1995) examining the concepts of intelligibility, accentedness and comprehensibility. Other research examining the constraints of the critical period and ultimate attainment in pronunciation conducted by James Flege and colleagues, as well as a number of other researchers (e.g., Moyer 2004) have rewritten the book on ultimate attainment in pronunciation acquisition, casting more attention on the importance of individual and social factors and less on biological ones.

The paradigm on teaching pronunciation, meanwhile, has also shifted away from the traditional emphasis on vowels and consonants to a prominent focus on prosody, the suprasegmentals of language. What this means is that the pronunciation research and teaching of today is very different than it was 40 years ago. Unfortunately, the stereotype that pronunciation means little more than endless drilling remains strong in many people's minds.

One of the consequences of the uneven professional infrastructure for training teachers (Murphy, 1997) is that teachers today feel more unprepared to teach pronunciation than in previous generations. Research across the English speaking world has been very consistent about this. Teachers are underprepared or uneasy about teaching pronunciation in the US (Morley, 1994), in Canada (Breitkreutz, Derwing & Rossiter, 2002), in the United Kingdom (Burgess & Spencer, 2000) and in Australia (MacDonald, 2002). The same lack of preparation has also been reported among teachers of Japanese (Kawai & Hirose, 2000), so it appears that the marginalization of pronunciation in the language curriculum is very much a worldwide phenomenon.

The infrastructure for professional preparation for pronunciation teaching is in disrepair, but it is not completely gone. Many influential teachers and researchers have been working diligently to carry out research, run training workshops and write books for teachers, provide materials for students in the classroom, migrate and create new materials for computer interfaces, and develop high quality, theoretically defensible courses in pronunciation as part of teacher training programs. Nonetheless, the infrastructure is not sufficiently available to allow pronunciation to take its deserved and equal role at the language teaching table.

Charles Swindoll, an American preacher, once gave a series of sermons about the biblical account of rebuilding the wall of Jerusalem around 500 BCE. The city had been destroyed, other

people had moved into the area, and the newcomers who returned from exile to rebuild the city faced obstacles on every side. Swindoll entitled the series, “Hand me another brick!” In much the same way, the first Pronunciation in Second Language Learning and Teaching conference is a modest attempt to place another brick or two on the infrastructure of a rebuilt approach to pronunciation by providing a professional space for teachers and researchers to meet together and discuss theory, trends and practice in pronunciation. The electronic proceedings of the conference are another brick, a place in which the wide variety of studies and approaches can be shared with a wider audience who may not easily have access to the wide variety of professional journals in which pronunciation-oriented research may be found. For readers of the proceedings, we hope that you will find much here that encourages you to join in the rebuilding, such that, in years to come, you can join in adding another brick or two to the rebuilding.

## PRESENTATIONS AND PROCEEDINGS

The first pronunciation conference was held jointly with the [Technology in Second Language Learning Conference](#) at Iowa State University and included approximately 90 participants from 10 US states and 5 foreign countries. [The portion of the conference schedule](#) devoted to pronunciation included two plenary addresses (from Wayne Dickerson and Tracey Derwing, both included in this volume). In addition, there were more than 20 other paper and poster presentations, 7 of which are in this volume. One presentation, entitled “Language awareness and second language pronunciation: a classroom study,” by Sara Kennedy and Pavel Trofimovich, is published in *Language Awareness* (2010, vol. 3).

## PLENARY ADDRESSES

Wayne Dickerson, from the University of Illinois at Urbana-Champaign, has over 35 years of experience in teaching pronunciation and training teachers during their MA in TESL studies. During his time at Illinois, he has influenced hundreds and hundreds of teachers. I took his Applied Morphology and Phonology class during my first semester in the ESL Master’s program, and I was hooked from the first day. The class was challenging, exciting and eye-opening, and the framework that he taught for teaching pronunciation and understanding the sound system of English continues to influence my thinking 25 years later. Wayne is a teacher’s researcher, always trying to see what learners need to become their own teachers, developing techniques to help make connections between the written word and how it is pronounced, and giving us the tools both to evaluate published materials and write our own.

His plenary address, *Walking the walk: Integrating the story of English phonology*, is a discussion of why and how he has changed the way he teaches teachers. In the plenary, he describes how our stated beliefs that phonology is an integrated system of creating meaning is at

---

odds with how we actually present the fundamental facts about the English sound system. Typically, we still teach the segments of language first, and then move on to suprasegmental features of stress, rhythm and intonation, teaching each as a separate subsystem that can be adequately addressed without reference to the other subsystems. In his paper, he discusses how this ordering actually gets us into trouble, and that we end up trying to hide the jams our teaching gets us into. (For example, we teach vowels and have to address [ə], the most numerous vowel in English. It is also a vowel sound that makes no sense unless we understand the rhythmic system that gives rise to it, creating a jam in which we can only explain the sound by making reference to a portion of the sound system that we have not yet addressed.) Then he goes on to discuss how a change he has made in his own course, addressing the sound system first from the rhythmic system helps avoid the difficulties that are inherent in the traditional way of presenting the sound system. The solution that he has integrated into his own teaching should be a challenge to others who teach teachers about the sound system of English and who wish to have their courses reflect the priorities we say we have about helping our students learn to pronounce English more effectively.

Tracey Derwing is Professor of TESL and the co-director of the Prairie Metropolis Centre for Research on Immigration, Integration and Diversity at the University of Alberta. Her name is familiar to everyone who has looked for research on the intelligibility of English and the effect of pronunciation on spoken intelligibility. With Murray Munro at Simon Fraser University in British Columbia, she has framed the research agenda for pronunciation since the early 1990s. Tracey started her career as an ESL teacher teaching pronunciation to immigrant students in Edmonton, Alberta, and her experiences as a teacher who had little to turn to except minimal pairs exercises have continued to inform her research agenda so that both researchers and teachers will find much that is useful in her many articles.

Her plenary, *Utopian Goals for Pronunciation Teaching*, starts from the recognition that pronunciation, despite the importance it plays in judgments of spoken intelligibility, continues to be a much neglected part of language teaching, much to the disservice both of teachers and of learners. She discusses nine action points for addressing this neglect: Changes to the way we educate ESL/EFL teachers, attention to integrating pronunciation into the language teaching curriculum, a greater focus on intelligibility rather than accent, increased pronunciation research, more appropriate uses of technology, enhancement of native speakers' listening skills, new attention to pronunciation in assessment, and strategies for increasing newcomer's opportunities to interact with native speakers. Far from being Utopian goals, she asserts that they should be seen more as a "to do list" and that each of the ideas are being, and can be, implemented now rather than waiting for some far-off future.

### **INTELLIGIBILITY, COMPREHENSIBILITY, AND ACCENTEDNESS**

Two papers in this collection directly address issues related to intelligibility, comprehensibility and accentedness (Munro & Derwing, 1995). Jennifer Rasmussen and Mary Zampini (Le Moyne



College) report on a study that addresses the improvement of listening skills for learners of Spanish. Most intelligibility research focus on how intelligible learners' speech is to native speakers of the target language, but an equally critical aspect of intelligibility is the learner's ability to understand native speech. Rasmussen and Zampini examine the development of listening ability for three features of Andalusian Spanish: Aspiration or deletion of syllable final /s/, synalepha, or linking/elision across word boundaries, and the pronunciation of the interdental fricative, /θ/. An experimental group of 2nd and 3rd year Spanish learners was trained to listen for these phonetic features in Andalusian Spanish, and their performance was compared with that of a control group. Results were mixed, with the experimental group showing significantly better improvement for one feature while the others showed no difference between the control and experimental groups.

The second paper, *Factors in Raters' Perceptions of Comprehensibility and Accentedness*, by Heesung Grace Jun and Jinrong Li (Iowa State University), employs verbal protocols (think alouds) to examine why three NS and three NNS raters judged NNS spoken performance as comprehensible. The researchers asked raters to first listen to the utterances and rate each for comprehensibility, followed by a think aloud session in which they discussed why they rated each sample as they did and what features they noticed that impacted their ratings. Following, this, the raters listened again and rated each sample for accentedness. Results showed that NS and NNS raters cited different features for their ratings, with NS raters focusing more heavily on global impressions and NNS raters citing specific pronunciation errors.

## **ACQUISITION AND ATTITUDES**

The three papers in the second section address the acquisition of pronunciation and learners' attitudes toward pronunciation. In the first paper, *The Effects of Self-Monitoring Strategy Use on the Pronunciation of Learners of English*, Sue Ingels (University of Illinois) examines a topic that is beginning to get an increasing amount of attention, the teaching of learning strategies for better learning of pronunciation. In the study, she looks at the effectiveness of training learners to monitor and correct their non-target use of English suprasegmentals using three different strategies or strategy combinations: Listening (L), Listening+Transcription (LT), and Listening+Transcription+Annotation (LTA). Using Listening alone appeared to help learners to monitor and improve, although the LT and LTA strategy combinations may have led to greater improvement for certain suprasegmental targets.

In the next paper, Fran Gulinello (Nasau Community College) reports on a longitudinal study of changes to the vowel systems of adult native Spanish speakers learning English. The study is a carefully constructed laboratory study examining 11 stressed, non-diphthong vowels spoken in comparable contexts. The findings show that speaker production of the vowels changed over time, including seeing two vowels merge into one category, one vowel splitting into two categories, and vowels shifting from one category to another. The paper argues that the interlanguage phonology of learners should be seen not only in light of its approximation to the

---

target language but also as a unique system in its own right. The change of the pronunciation of one vowel in a system cannot be seen as isolated from changes in other vowels, and the changes in one vowel should affect how we understand and teach the pronunciation of the new L2 vowel system.

The last paper in this section, *ESL Learners' Attitudes toward Pronunciation Instruction and Varieties of English*, comes from Okim Kang (Northern Arizona University). In it, she examines learners' expectations toward learning pronunciation and their attitudes toward the accents of different inner circle varieties of English, specifically in New Zealand and the United States. Her findings showed that students studying in New Zealand far more likely to be dissatisfied with the pronunciation instruction they received than were students studying in the United States. She also found that those in New Zealand were could be more ambivalent or even quite negative toward the variety of English they heard in comparison with the learners who studied in the United States. She suggests that learner attitudes be taken into account to provide better pronunciation, and suggests that the differences shown in response to these two varieties have implication far beyond these two settings to the teaching of pronunciation in outer and expanding circle contexts.

## **DESCRIPTIVE RESEARCH**

The last section of the proceedings includes two papers that are descriptive in nature: one focusing on the pronunciation of the past tense (-ed) morpheme by Somali learners of English, and the other examining the use of sentence focus in authentic materials. In the first paper, Ettien Koffi (St. Cloud State University), examines why Somali learners of English (a significant immigrant population in Minnesota) consistently have trouble pronouncing past tense verbs correctly in certain contexts even when they pronounce it perfectly in others. He analyzes the syllable structure of Somali and compares it to English, showing how both the insertion of epenthetic schwa in verbs like *kissed/jumped/kicked* and the deletion of the suffix in verbs like *begged* can be understood by understanding both the syllable structures of English and Somali. Koffi's paper is an excellent example of how the use of linguistic knowledge and reasoning is so critical to helping teachers better address their learner's needs.

In the next paper, *Authentic speech and teaching sentence focus*, Greta Muller Levis and John Levis (Iowa State University) examined how authentic speech can be used to create teaching materials for sentence focus (e.g., How ARE you? FINE. How are YOU?). Focus is an essential part of communicating pragmatic meaning in English, and it is a suprasegmental feature that is prominently displayed in most published teaching materials. Since focus is typically connected to the information structure of discourse, the teaching of focus is also tied closely to highlighting new information new information. Most published materials, however, are constructed rather than authentic. The paper examines the issues involved in creating teaching materials for non-final new information from authentic academic and non-academic discourse. Results suggest that texts with multiple examples of non-final new information are

rare, calling for adaptation by teachers, and that most reasonably useful authentic texts also include uses of focus that are difficult to explain by using reference to new and given information alone. The paper ends with suggestions for using authentic and adapted materials to teach sentence focus.

## REFERENCES

- Breitkreutz, J., Derwing, T.M., & Rossiter, M.J. (2002). Pronunciation teaching practices in Canada. *TESL Canada Journal*, 19, 51 – 61.
- Burgess, J. & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28, 191-215.
- Deng, J., Holtby, A., Howden-Weaver, L., Nessim, L., Nicholas, B., Nickle, K., Pannekoek, C., Stephan, S., & Sun, M. (2009). English pronunciation research: The neglected orphan of second language acquisition studies. *Prairie Metropolis Centre Working Paper Series*, WP05-09, Edmonton, AB.
- Hinofotis, F. & Bailey, K. (1981). American undergraduates' reactions to the communication skills of foreign teaching assistants, in J. Fisher, M. Clarke and J. Schachter, (Eds.). *On TESOL '80 : building bridges : research and practice in teaching English as a second language : selected papers from the fourteenth annual convention of Teachers of English to Speakers of Other Languages*. Washington, DC: TESOL.
- Fotos, S. (2001). Cognitive approaches to grammar instruction. In M. Celce-Murcia (Ed.), *Teaching English as second or foreign language, 2nd ed.* (pp. 267-284). Boston: Heinle & Heinle.
- Gilbert, J. (2010). Pronunciation as orphan: what can be done? *SpeakOut: The newsletter of the IATEFL pronunciation special interest group*, 43, 3-7.
- Kawai, G & Hirose, K. (2000). Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication*, 30, 131-143.
- Kennedy, S. & Trofimovich, P. (in press). Language awareness and second language pronunciation: A classroom study. *Language Awareness*, 19(3).
- Krashen, S. (1985). *The input hypothesis: issues and implications*. London: Longman.
- Larsen-Freeman, D. (2001). Teaching grammar. In M. Celce-Murcia (Ed.), *Teaching English as second or foreign language, 2nd ed.* (pp. 251-266). Boston: Heinle & Heinle.
-

- Long, M. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly* 24, 649–66.
- MacDonald, S. (2002). Pronunciation—views and practices of reluctant teachers. *Prospect*, 17 (3), 3–18.
- Moyer, A. (2004). *Age, accent and experience in second language acquisition*. Clevedon: Multilingual Matters.
- Munro, M. & Derwing, T. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Murphy, J. (1997). Phonology courses offered by MATESOL programs in the United States. *TESOL Quarterly*, 31(4), pp. 741-764.
- Scovel, T. (2000). A critical review of critical period research. *Annual Review of Applied Linguistics*, 20: 213-223
- Vanpatten, B. (2004). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum.

## **Walking the Walk: Integrating the Story of English Phonology**

**Wayne B. Dickerson**

University of Illinois at Urbana-Champaign

We pedagogical phonologists talk the talk: We assure our students of what we believe, namely, that all the parts of English phonology make up a single, intricate system of subsystems. The problem is that neither our descriptive nor our pedagogical texts present English phonology this way. Furthermore, we have not always taught it this way. We haven't walked the walk.

Recently the challenge of moving my pedagogical phonology course to a hybrid—in-class and online—format gave me the opportunity to try to practice what I preach. This paper describes how I have come to retell the story of English phonology in a way that is at least truer to its integrated nature than the way I told it before. Walking the walk has been a journey of several years. In retracing some of my steps, I want to show what can be done for our students. The result has been satisfying—a course that is new and interesting even to those who have some background in phonology. But most importantly, it is a course that does a better job now than before to prepare pronunciation teachers, materials developers, and those who want to go on with their study of theoretical or applied phonology.

### **THE PRESENTATION OF PHONOLOGY**

ESL teachers who have had formal TESL preparation have usually been exposed to English phonology through an introduction to linguistics or a descriptive phonology course. If they were lucky, they took a pedagogical phonology course. However they became acquainted with the English sound system, the outline of content no doubt followed the typical sequence, beginning with segmental phonology and then moving to suprasegmental phonology, e.g. Fromkin, Rodman, & Hyams (2003); Roca & Johnson (1999).

Until recently, my English phonology and morphology course for TESL teachers looked like that, too, as illustrated by the following list of topics.

---

Segmentals	Suprasegmentals
Consonant Phonemes	Word Stress
Vowel Phonemes	Phrase Rhythm
Spelling	Construction Stress
Variability	Primary Stress
Consonant Phonetics	Intonation
Vowel Phonetics	

Like all phonologists, whether in textbooks or in classes, I professed the belief that all of these subsystems of phonology are woven into a single system. It is a truth that we hold to be self-evident. On the other hand, as a matter of practicality, to study the system or to present it, it seems that our default strategy has been to focus on each subsystem in itself. That's why I had a unit on consonants, another on vowels, one on word stress, another on construction stress, and so on.

### Issues with the typical presentation

Not long ago I began to think about the discrepancy between what we profess and what we actually do—claiming phonology to be a whole fabric but giving attention only to its threads. We become so absorbed in the intricacies of each thread—because they are so fascinating in themselves—that we have not gotten around to showing how the threads are woven into this amazing whole we call the sound system. As an article of faith, we know it to be so. We talked the talk of integrated subsystems, but we rarely walked the walk of consistently demonstrating that integration.

Before we go too far, perhaps we should ask ourselves: What's so wrong with this segmental vs suprasegmental organization? After all, its popularity seems to say that it works pretty well. It works, but not well. First and foremost, this organization does not represent the reality of the sound system. The reality is that these components interlock. To ignore this point, as many of us have done, is to misrepresent the facts. Second, since we tend to teach pronunciation the way we were taught phonology, we tend not to teach pronunciation in a way that helps ESL learners integrate the sound-system components they are learning. We may teach a little of this and a little of that in every pronunciation lesson, but do we present the pieces so that they build on each other?

For instance, we learned the unstressed Schwa /ə/ as a vowel on the vowel chart. So how many of us have taught /ə/ as another vowel sound? I did. But /ə/ is not like any other vowel sound; it's

a byproduct of rhythm, unlike the other vowels. If we don't teach /ə/ as it relates to rhythm, learners have no way to understand why it's so important, why it's three times more frequent in running speech than any other vowel sound (Woods, 1983, p. 50), or how to use it to create rhythm. Schwa is not a stand-alone vowel; it's part of larger system. This is what I mean about teaching the parts of the system in a way that shows their relationship to each other.

A third problem with focusing on sound-system components and ignoring their interactions, is that this approach to the sound system gets us into serious jams both in presenting the facts of phonology and in teaching pronunciation. What's a jam? It's when you're trying to describe a phenomenon in phonology, and you find that you can't fully describe it without reference to some other part of the sound system that you haven't yet introduced. It's like teaching /ə/ as one of the English vowels without teaching its rhythmic context. That's a jam.

I want to illustrate two more jams in some detail. One comes from the segmental side of phonology; the other, from the suprasegmental side. Then we'll talk about how to avoid all jams and also convey the reality of interdependent components.

### **Jam #1: Vowel lengthening**

Most of us are familiar with the description of vowel lengthening—a topic from the segmental part of our content—specifically vowel phonetics. You've read the claim in pronunciation books that vowels have longer duration when they come at word ends, as in the first row below, or when they are followed by a voiced consonant, as in the second row. When followed by a voiceless consonant, on the other hand, the vowel is much shorter, as in the third row below.

1.	say	go	plea	try	crew
2	save	goad	plead	tribe	cruise
3	sake	ghost	pleat	trike	croup

This description is wrong in so far as it is incomplete. If learners actually lengthened all vowels in open syllables (first row) and before voiced consonants (second row), their speech would sound hideous. Why? Because that's not what English speakers actually do. As it turns out, vowel lengthening is as much a function of primary stress—a suprasegmental—as it is a function of its segmental environment. Without primary stress, vowels in different environments are hardly different at all. To illustrate, say these two sentences aloud. Which plead is shorter in duration?

*How did he plead? I'll plead my case.*

The shorter *plead* is in the second sentence, *I'll plead my case*.

---

The two *pleads* have the same segmental environment in both sentences—a vowel before a voiced consonant; the difference is that the two *pleads* are in different suprasegmental environments. The *plead* in *How did he plead?* carries primary stress, and the other does not.<sup>1</sup> So why do we hear a difference between pairs of words like *save* – *sake* and between *goad* – *ghost*? The reason is that we say each word as a single phrase with its own primary phrase stress. We introduce primary stress without realizing it.

My point is this: We should not describe or teach vowel lengthening solely as a segmental phenomenon because it isn't; it's critically dependent on the suprasegmental of primary stress. You can see the jam we get ourselves into—and the misinformation we convey—when we teach vowel lengthening without reference to primary stress.

### **Jam #2: Compound number stress**

One more example of a jam before we talk about solutions. We all know about constructions like compound nouns, compound adjectives, compound numbers, compound verbs, and multiword or phrasal verbs. The stress of at least some constructions is typically considered in textbooks—in phonology and pronunciation textbooks alike—right after word stress and before primary stress—all suprasegmentals. We may learn the word stress rule for *elevator* and *operator* and then the construction stress rule that accounts for the stress of the compound noun *elevator operator*. And then we see the implication of this rule because it guides the primary stress placement in a sentence like:

*At the Ritz, they still have élevator operators.*

The order of word stress before construction stress before sentence stress sounds reasonable. If it didn't, it wouldn't be done this way.

But it's not that simple—for any of the constructions, actually. As an example, let's look at the stress of compound numbers like *fourteen*, *twenty-four*. Read each of these sentences, and listen for the position of the heavier stress in each compound number.

*I counted to fourteen.*                      It's on *-teen*.

*She was twenty-four at the time.*      It's on *-four*.

So the first rule about the stress of compound numbers identifies the default stress of compound numbers:

Rule 1:            If the compound number does not modify a noun, then its heaviest stress falls on the last element of the compound: fourEEN.

Now say this sentence with the primary stress is on *planes*. Which part of the compound number has the heavier stress?



*I counted fourteen planes.*

It's on *four-*.

The second rule takes into account that the compound number modifies a noun:

Rule 2: If the compound number modifies a noun then its heaviest stress shifts to the first element of the compound: FOURteen.

So far these rules match what we find in most pronunciation textbooks that acknowledge stress changes in compound numbers. But that's only part of the story. In the sentence above, planes carries the primary stress. Things are different when fourteen carries the primary stress. For example, let's say I've been talking with my wife about the long row of planes lined up on the tarmac waiting for the go-ahead from the tower. Studying the situation, she might say,

*I counted fourteen planes there.*

Stress is on *-teen*.

The word *planes* in this conversation is old information because we've already been talking about them. The new information is *fourteen*<sub>2</sub> and that is where the primary stress falls. The stress of *fourteen* is the default, second-element stress as described in Rule 1. There is no stress shift despite a following modified noun.

So the stress of compound numbers isn't just about whether or not the number modifies a noun—a phrase-level observation; it is also about whether or not the compound number is primary stressed. And that is a discourse-level decision. This is why rules 1 and 2 have to be revised.

Rule 1 (revised) If the compound number carries the primary stress, or if it does not modify a noun, then its heaviest stress is on the last element of the compound: fourEEN.

Rule 2 (revised) If the compound number does not carry the primary stress but modifies a noun, then its heaviest stress shifts to the first element of the compound: FOURteen.

My point is this: It looks reasonable, in terms of the size of the units, to talk first about word stress, then about construction stress, and finally about primary stress. But clearly, the discourse-level decision about the position of the primary stress must come before construction stress, or we cannot understand fully how compound number stress works.<sup>2</sup> We're in a jam if we try to reverse the order.

### **Jams: The consequence of linear organization**

Jams are all over the place in the traditional presentation. We get into a jam if we try to present aspiration before word stress. We find ourselves in a jam if we teach word stress before phrase rhythm. I've mentioned the jam of presenting /ə/ before phrase rhythm and detailed the jam of

---

teaching vowel lengthening without reference to primary stress. And you've seen the jam of presenting compound number stress before primary stress. There are many other jams, as well.

Jams illustrate this property of English phonology: It is not organized linearly. One subsystem provides the conditions for features in another subsystem to appear. And that subsystem creates the environment for features in the first subsystem. The interdependency of subsystems is in the nature of phonologies. And English is no exception. To reflect English phonology faithfully, our presentation of it should highlight, not downplay, the interdependencies.

### ESCAPING THE DILEMMA: DOMINANT THEMES

So how do we get out of this dilemma when there is no ideal order of subsystems? First of all, we need to understand the dominant themes of English phonology, those that have the greatest impact on the rest of phonology. This helps with an initial organization of topics in phonology so that many of the interdependencies are accounted for from the start. Then, where the reorganization of topics by dominant themes does not help, we need to have a strategy for mixing information from the different subsystems in just the right amount, at just the right moment, so that all relevant information is in place with which to describe any phenomenon.

So let's talk about dominant themes. As far as I can tell, there are only two in all of phonology, and virtually everything in phonology depends on them. Both are oriented toward communication. That shouldn't be a surprise, given the purpose of oral language.

First, there is phrase rhythm and semantic prominence. We all know that in English-style rhythm, there is a contrast between peaks and valleys of stress. Peaks of stress make vowels louder, longer, and higher-pitched. Those contrast with valleys of stress with their quieter, shorter, and lower-pitched vowels. We can think of this as the vertical dimension of rhythm (Figure 1).



Figure 1. The vertical dimension of English rhythm: Peak-valley contrasts.

This 'bumpiness' is not without purpose. It turns out that the words in the peaks are more semantically important than the words in the valleys. Stress is iconic. That is, it mirrors the importance of words for communication. One syllable of semantically prominent words like content words and loud function words (question words, negative words, and demonstrative pronouns) carries greater stress than any of the semantically peripheral words like soft function words. This semantic link to phrase rhythm helps us catch and process messages quickly. We

have come to depend on that link for understanding every phrase. To accommodate that dependency when we speak, moment by moment we are either making words stand out from the background or causing them to recede into the background. That is, in every phrase we are constantly modulating our word stress—from greater to lesser—and our vowel quality—from full vowels to reduced—to highlight the semantically more important words and obscure the semantically less important ones. This is an exceedingly difficult challenge for learners.

Outside of this rhythm-semantics connection, there is no way to explain the valley stress of soft function words in English. Why don't they have greater stress? Because their relative unimportance to the message does not permit it.

If peak-valley contrasts are the vertical dimension of rhythm, then the timing of peaks is its horizontal aspect (Figure 2). While we do not create perfect regularity, we do make a serious effort to have our peaks appear with some regularity. To do this we squeeze valley syllables so they don't take up more time than is necessary. Squeezing valley syllables is what makes possible the regularity we hear in the appearance of peaks.

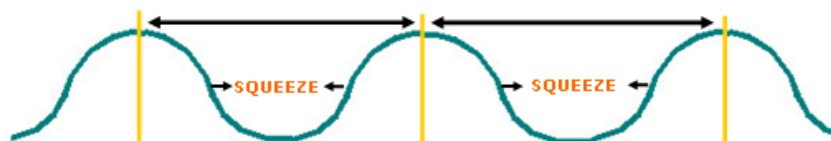


Figure 2. The horizontal dimension of English rhythm: Peak-to-peak timing.

How do we squeeze? If you look at the consonant and vowel phonetics presented in a typical phonology course, except for a few features like nasalization, epenthetic stops, and clear and dark el, nearly all of the phonetic features are involved in squeezing syllables. Either the phonetic environment provides the conditions for trimming sounds like [t] and [d] at the ends of clusters, like vowels and consonants in contractions, like the vowel we lose when we say *family* as two syllables instead of three, like the loss of initial *h* of *he, him, his, her, have, has, had*, etc. in valleys. Or phonetic phenomena are, themselves, cases of squeezing, like vowel reduction, flapping (*The water's bitter*), or syllabic consonants (*It's a model garden*). All of these phonetic devices eliminate articulation time in valleys and thereby promote regular rhythmic timing. Again, which syllables receive this crushing treatment? Mainly the syllables of semantically unimportant words, and the non-peak syllables of other words.

While we used to look at phonetics primarily in terms of the allophonic members of consonant and vowel phonemes, that is not the perspective I now take in my pedagogical phonology course. Instead, we look at phonetics in terms of its primary function, namely, to systematically squeeze

---

time out of rhythmic valleys by all means possible. To make that point, the suprasegmental topic of phrase rhythm comes before phonetics in our syllabus.

So phrase rhythm and semantic prominence shape virtually everything that happens within the phrase. That's the first dominant theme.

And the rest? The rest of phonology has to do with discourse and is dominated, not by rhythm, but by discourse meaning. What do I mean by discourse meaning? We've already referred to the role of primary stress to highlight new information; that's discourse meaning. How do we tell each other that something is old information? We destress it. That signals discourse meaning. We also use primary stress to contrast one thing with another. That's discourse meaning, too. How do we communicate an aside—*He got there first, the jerk?* We drop our voice on *the jerk* and use a special intonation which tells the listener our opinion of what we've just said. And if we want to indicate that we're uncertain about something or even to imply the opposite of what we're saying? We use a particular intonation—*He's a good teacher* √. This is discourse meaning, too. These are all meanings that we communicate to each other with the resources of stress, pitch, and intonation.

The importance of finding these dominant themes is that it has reorganized the content of my pedagogical phonology course in significant ways. Here's a hint of what the reorganization looks like (Figure 3).

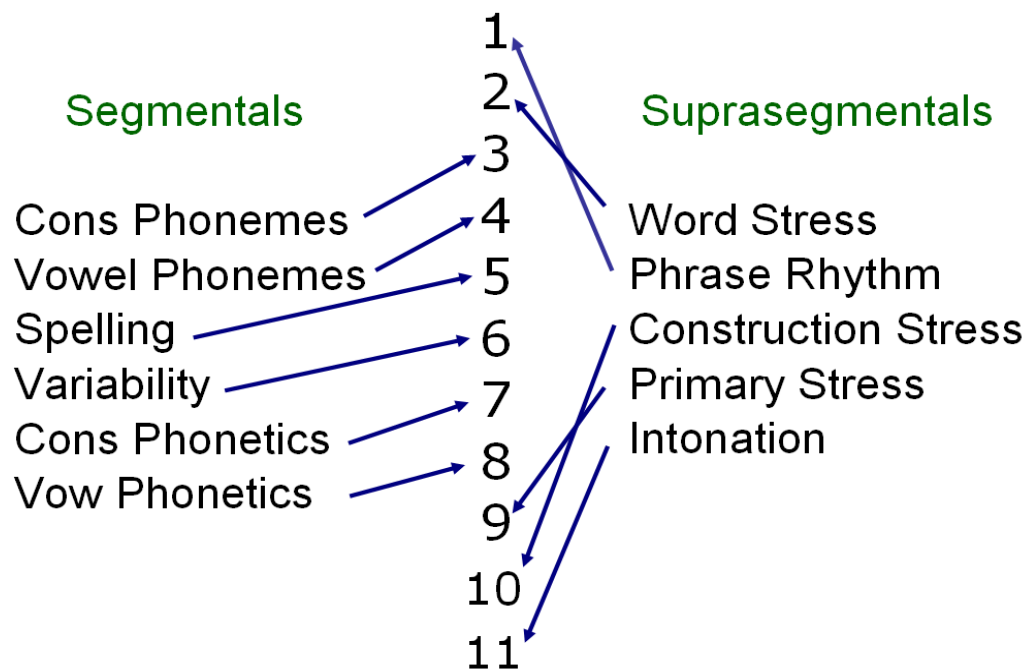


Figure 3. Course reorganization to integrate segmental and suprasegmental topics.

The old course topics are listed—segmentals on the left and suprasegmentals on the right. Down the middle is the numbered list of topics in the reorganized course. By starting with phrase rhythm, word stress makes sense; rhythm and word stress before full and reduced vowels help students understand how these vowel qualities promote peak-valley contrasts. And by starting with rhythm, virtually all of phonetics makes sense because nearly all of it is there to facilitate rhythmic timing.

So the first step in solving the problem of interdependent subsystems is to build our phonology course around its dominant themes. By getting the fundamentals right first, relationships among subsystems are more properly aligned from the start. When we do that, you can see that segmental and suprasegmental subsystems are beginning to integrate with each other. Segmental topics make more sense when introduced with the suprasegmental topics of phrase rhythm and word stress.

But where does construction stress go? No matter where it goes, it is clear that by teaching phrase rhythm before construction stress, we can understand both *seventy-FIVE* and the stress-shifted *SEventy five SEConds* in rhythmic terms, not simply as stress patterns. Looked at through the rhythmic lens, we can easily see that the contrast of valley-peak in *seventy-FIVE* and the contrast of peak-valley-peak in *SEventy-five SEConds* are rhythmic effects. So the order of rhythm before construction stress works.

But, as we also saw, phrase rhythm alone cannot explain why we go back to the default valley-peak stress in this dialog about the shuttle launch:

A<sup>1</sup>. We're counting the seconds to láunch.

B. Where's the count nów?

A<sup>2</sup>. We paused at *seventy-five* seconds,

but expect to resume shórtly.

The explanation rests with discourse meaning, and the placement of the primary stress. When *seconds* in A<sup>2</sup> is old information, pointing back to the previous reference in A<sup>1</sup>, *seconds* is moved to a valley because old information is destressed. To its left, *seventy-five* is the primary-stressed new information and, according to Rule 1 (revised), will carry a peak stress on *five*.

The full pattern of compound number stress is available only when the discourse topic of primary stress has been covered. When we study all the other kinds of construction stress—that of compound nouns, compound adjectives, compound verbs, and phrasal verbs—it turns out that they, too, depend on discourse notions related to primary stress. Given that they all need discourse stress information to fully explain what is going on, it makes good sense to move construction stress to a point after primary stress, as shown in Figure 3. In that way, when the

---

primary stress information is needed for construction stress, it is there. This reordering helped us out of multiple jams.

### **ESCAPING THE DILEMMA: JIT PRINCIPLE**

Getting the themes right makes a huge difference in the ordering of topics. But since subsystems are more interdependent than this, no single ordering of subsystems solves all problems. We must take the next step which is to understand how to mix the subsystems even more. That will allow us to have the right information available when we need it to make sense of phenomena that are still not accounted for by reorganizing according to dominant themes.

To take the second step, we need the help of Toyota. You may have heard of Toyota's dilemma early in its life as a car-maker. In tiny Japan, there was not space enough for huge storage buildings for inventory, and car lots the size of football fields for finished product. So in the 1950s Toyota executives came to the States looking for a different way to make cars. Here they found Piggly-Wiggly grocery stores—America's first self-service grocery store, founded in 1916 in Memphis, Tennessee.

What interested Toyota about Piggly-Wiggly stores was that they carried almost no inventory. Everything they had was on the shelves. When they began to run low on items, a shipment was already on its way to fill the shelves. They represented the most thorough implementation of the just-in-time inventory principle—JIT, for short. They kept tabs on everything. They put in orders for just the right things, to be delivered at just the right time in just the right amounts to the right grocery stores so that grocery stores didn't have to have big storerooms. As a result, they could devote more space to customer-accessible isles. Of course, this is the model at the heart of all big chain stores now—Walmart, Target, Lowes, Office Depot, and on and on.

This is what Toyota was looking for—a system on the input side that would supply car parts in just the right quantities at just the right time to just the right assembly plant so that plants would not have to store car parts in inventory. And on the output side, they made cars to order so that they didn't have to store them on huge surface lots. When new cars were produced, they were shipped.

This JIT principle is also the key to presenting phonology as an integration of subsystems. To illustrate, let's return to our example of vowel lengthening. We cannot present the whole story of vowel lengthening without information from segmentals and information from suprasegmentals, namely, about the location of the primary stress in a phrase. Even in our reorganization, primary stress still comes much later than vowel phonetics. This is the reason for the JIT principle.

All we need is the right information at the right place at the right time. Specifically, to tell the story of vowel lengthening, we just need to know where primary stress typically falls in a phrase. That is, we need just this fact: Primary stress typically falls on the last content word in a phrase. This tidbit certainly does not tell the whole story of primary stress placement. But that's not the

point. The point is to have at hand just what is necessary to present the notion of vowel lengthening faithfully. When we do come to the actual presentation of the primary stress system, we can start with the basic notion and elaborate from there. No harm done.

The JIT introduction of just the right information in just the right amount at just the right place takes only a brief amount of time. But it makes it possible not only to tell the whole story about a phenomenon, but if we are explicit about what we are doing when we insert information from another component, it also convincingly demonstrates the interrelatedness of subsystems.

## **CONCLUSION**

When we look at pedagogical English phonology from the perspective of dominant themes and then identify all the places where subsystems intersect and JIT is needed, we come out with a different kind of syllabus from before.<sup>3,4</sup>

And what is the effect of this reorganization? I have been surprised by how much more interesting the details of phonology have become to MATESL students. The sound system is no longer a long list of components. Now they understand how the disparate parts of phonology interconnect for a purpose, namely to implement the dominant themes. And from this better understanding of phonology has come a clearer conception of what ESL learners need to learn—what themes need to be emphasized in the pronunciation course. As a direct consequence, the materials MATESL candidates develop reflect that better understanding. And my students' ESL students are exposed to a more holistic organization of pronunciation content in which the details fit into larger schemes. To say the least, this is a very gratifying set of outcomes.

In summary, we have long talked the talk of integrated subsystems of phonology because we believe it, as do all phonologists. But it wasn't until the last couple of years that I have been able to walk the walk and actually demonstrate the integration by showing how pieces from different subsystems work together to give us the sound system phenomena we have.

To do this has required rethinking English phonology in two particular ways. In the first place, it has taken a thorough reexamination of the fundamentals of English phonology to identify the dominant themes that are most responsible for regulating how we speak. Those are phrase rhythm as linked to semantic prominence, and discourse meaning. In the second place, it has taken an understanding of the JIT principle to jar us loose from a subsystem-by-subsystem presentation of phonology and to allow us to introduce just the right information, in just the right amount, at just the right time, and in just the right places to tell the integrated story of how our sound system works.

---

## ABOUT THE AUTHOR

Wayne Dickerson is Professor of Linguistics at the University of Illinois at Urbana-Champaign where he teaches pedagogical phonology in the MATESL program and ESL pronunciation for international students. He is certified as a Master Online Teacher and teaches the pedagogical phonology course online, as well. An applied phonologist, his principal interest is in the application of technical linguistic insights to practical teaching tasks, with an emphasis on English stress, rhythm, orthography, and materials development. His most recent books are *Stress in the Speech Stream: The Rhythm of Spoken English* (1989), published by University of Illinois Press, and (with co-author Laura D. Hahn) *Speechcraft: Discourse Pronunciation for Advanced Learners* (1999) and two workbooks, published by The University of Michigan Press.

## NOTES

<sup>1</sup> Besides the phonetic environment—before a voiced consonant—and the suprasegmental environment—under primary stress—the *plead* in *How did he plead?* also satisfies a third requirement. The vowel is in the last syllable of the phrase. Having to meet all three requirements simultaneously is why lengthened vowels do not occur with high frequency in discourse.

<sup>2</sup> An ESL pronunciation textbook that presents compound number stress according to the revised rules above is Hahn & Dickerson (1999a, b, c).

<sup>3</sup>The conference handout included the following course outline covering the descriptive phonological content of the course.



<p>I. Phrase-level phonology: Phrase rhythm and semantic prominence</p> <p>A. Peak-valley contrasts of rhythm</p> <p>B. Word stress: ^ ` ~ (´ delayed until primary stress is presented)</p> <p>C. Consonant phonemes</p> <p>D. Syllable structure and clusters</p> <p>E. Vowel phonemes (unstressed before stressed)</p> <p>F. Timing of rhythm</p> <p>G. Natural speech phenomena</p> <p>H. Consonant and vowel phonetics</p> <p>I. Phonological processes (palatalization, voice assimilation, {Z} and {D} morphemes)</p> <p>J. Value of orthography (predicting word stress, segmentals, variability)</p>	<p>II. Discourse-level phonology: Discourse meaning</p> <p>A. Primary stress: ´ (subsystems: parenthetical, contrast, including emphasis, pragmatic, new-information)</p> <p>B. Construction stress (compound nouns, compound numbers, multi-word verbs)</p> <p>C. Intonation (functions, forms, meaning of intonation)</p>
---	---

<sup>4</sup>Practical work on teaching pronunciation is not outlined in note 3. The practical side of the course includes projects on analyzing learners' production, prioritizing ESL learner effort, designing pedagogical materials, learning teaching techniques for tutoring and classroom instruction, doing the actual tutoring (two 30-minute sessions), and finally reflecting on the tutoring experience. We also give students instruction on how to use various technologies for instruction: recording and editing of audio materials with Audacity and synchronous tutoring using Skype coupled with GoogleDocs for instantaneous written communication between tutor and tutee.

## REFERENCES

- Fromkin, V., Rodman, R., & Hyams, N. (2003). *An introduction to language* (7th ed.). Boston: Thomson/ Heinle.
- Hahn, L., & Dickerson, W. (1999a). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor, MI: University of Michigan Press.
- Hahn, L., & Dickerson, W. (1999b). *Speechcraft: Workbook for academic discourse*. Ann Arbor, MI: University of Michigan Press.
- Hahn, L., & Dickerson, W. (1999c). *Speechcraft: Workbook for international TA discourse*. Ann Arbor, MI: University of Michigan Press.
- Roca, I., & Johnson, W. (1999). *A course in phonology*. Oxford: Blackwell.
- Woods, H. (1983). *Vowel dimensions, length, stress and unstress*. Ottawa, Canada: Canadian Government Publishing Services.

## Utopian Goals for Pronunciation Teaching

Tracey M. Derwing

University of Alberta

As has often been noted in recent years, pronunciation instruction has received short shrift from researchers and teachers alike. Although there is a small and committed body of individuals who have worked to encourage the incorporation of pronunciation instruction in English as a second language (ESL) classes, pronunciation still tends to be the neglected component of many language programs. In this talk, some idealistic goals for pronunciation instruction will be laid out, and suggestions for how the TESOL community might work towards achieving those goals will be addressed. Changes to teacher education, increased pronunciation research, optimal use of technology, enhancement of listeners' skills, and strategies for increasing students' opportunities to interact with native speakers are identified as potential approaches to meeting students' communication needs.

### INTRODUCTION

In recent years several researchers and practitioners have pointed out that pronunciation seems to be the orphan of second language research and teaching (Derwing & Munro, 2005). It tends to be neglected in the second language (L2) classroom, and L2 teachers are somewhat intimidated by the idea of teaching pronunciation (Burgess & Spencer, 2000). However, pronunciation, both segments and prosodic factors (suprasegmentals) have been the subject of study for a very long time, as has pronunciation instruction. In 1665, back in the day of long book titles, Owen Price, master of arts and professor of the art of pedagogy, wrote a volume entitled *The vocal organ, or the art of teaching the English orthography, instruments of pronunciation, and the difference between words of like sound whereby any outlander, or mere English man, woman or childe may speedily attend to the exact spelling, reading, writing or pronouncing of any word in the English tongue without the advantage of its fountains, the Greek and the Latin*. Price concentrated primarily on the study of segmentals of English, but in 1787, suprasegmentals were the focus of attention in Walker's book, *The melody of speaking delineated, or elocution taught like music by visible signs, adapted to the tones, inflexions and variations of voice in reading and speaking with directions for modulation and expressing the passions*.

In 1904, Otto Jespersen wrote *How to teach a foreign language*, which was reprinted for the next 50 years. In his manual, Jespersen took language teachers to task for being afraid of the phonetic alphabet and he argued that "The use of phonetics and phonetical transcription in the teaching of modern languages must be considered as one of the most important advances in modern pedagogy, because it ensures both considerable facilitation and an exceedingly large gain in exactness. But these means must be employed immediately from the very beginning" (p. 170). He went on to say that "Just as easy as it is to get a good pronunciation in this way, just as

---

difficult is it to root out the bad habits which may become inveterate during a very short period of instruction according to a wrong or antiquated method” (p. 176). With the advent of the International Phonetics Association’s Alphabet (IPA), Jespersen, along with others, thought that pronunciation of a second language could be scientifically explained and improved. Roughly 50 years later, devoted teacher-educator, Earl Stevick (1957) made some key points with regard to pronunciation teaching: start early – accuracy matters; start big – focus on pitch, stress and rhythm; be consistent; spread your work – 4 sessions of ten minutes are better than 1 session of 60 minutes; teach in terms of contrasts; and practice with connected speech. He called his general approach the Oral Approach – and it had the same basic principles of Audiolingualism, including a strong emphasis on pronunciation and getting it right from the start. Both these methods stressed the importance of good oral productions.

Another method that emphasized the importance of pronunciation was the Silent Way, in which L2 students’ exposure to vocabulary was extremely limited in the first month. All their words were represented in wall charts and each letter was colour-coded to provide a visual representation of sound and spelling correspondences. The Silent Way, in its pure form, was not practiced in very many locations because it required considerable training on the part of the teachers. But its founder, Caleb Gattegno (1976), maintained that the method was highly successful in producing L2 speakers who had excellent pronunciation.

It is somewhat ironic that there could be an approach to teaching pronunciation that emphasized silence on the part of the teacher, but it is similarly puzzling that the communicative approach, which became widespread in the 1980s and is still very influential, would have so little to say about accent. ESL instructors who learned to teach using the communicative approach had little guidance when faced with students whose speech was almost completely unintelligible. There were some materials available, primarily minimal pairs contrasts such as Nilsen and Nilsen (1971), which were thought to help speakers of other languages to improve their productions. All conceivable contrasts that students might have difficulty with were listed in Nilsen and Nilsen’s volume, even the contrast between voiced and voiceless TH, despite the fact that practically no one confuses these two sounds. It is far more likely that speakers would substitute a ‘t’ or an ‘s’ for theta and a ‘d’ or a ‘z’ for thorn. There was a general consensus in the 1980s among many teachers that pronunciation instruction was ineffective, and that the only activity one could employ was repetition. It is not altogether surprising that this skepticism existed. In the first place, very few ESL instructors at that time had any TESL or linguistics background. There was limited access to good materials, with a few exceptions such as *Jazz Chants* (Graham, 1978). The only available technology was the language master machine, which could read computer cards. A student would record a sentence and feed it into the machine to compare his or her productions with those of a model.

## UTOPIAN GOALS

It was during this period, the early 1980s, that significant numbers of Vietnamese speakers arrived at the school where I taught ESL. Several students had an excellent grasp of English syntax and vocabulary but had great difficulty making themselves understood when they spoke. The phonology of Vietnamese, a tone language, differs dramatically from that of English. For example, in Vietnamese there are only six contrasting final consonants and no consonant clusters, compared to the over 200 word final consonants and clusters in English (Hultzén, 1965). My fellow ESL teachers and I learned how to teach pronunciation by trial and error; and, although our students ultimately benefited, it wasn't an ideal situation. This raises the question of what would be ideal. What would effective and efficient pronunciation teaching for L2 learners look like in a Utopia? The following nine characteristics would surely be included: increased attention from researchers; a focus on teacher education; appropriate curriculum choices; improved assessment; focus on intelligibility/comprehensibility; more useful software and other technology; a focus on enhancing native speakers' listening; no scapegoating of accents; and better strategies for integrating newcomers into the community. Let us go through these goals for our field one by one.

### 1. Increased Pronunciation Research

There are competing views as to the usefulness of applied linguistics research to the language classroom teacher, but when it comes to pronunciation, I am firmly of the belief that such research can be valuable. Take, for example, Hahn's (2004) study, which showed that primary stress makes a difference in how well people can understand utterances. This is helpful to know. However, how much attention does pronunciation get from second language acquisition (SLA) researchers? Adam Brown (1991) surveyed four journals between 1975-1988 and found that very few articles on pronunciation were published during that time. There is still a very small percentage of articles devoted to our field in the general ESL/SLA research journals, with a range of 2.7% to 7.4% from 1999-2008 (Deng et al., 2009). Some would argue that research isn't that useful; indeed recently there was a comment on a pronunciation listserv from an experienced practitioner that was quite disparaging of research, and which suggested that intuition is all that is necessary to design activities that will meet students' needs. Unfortunately, that isn't the case. Although there may be some individuals who are capable of determining what is best for the students and then implementing appropriate classroom procedures, more people are likely to avoid dealing with pronunciation altogether. Studies by Breikreutz, Derwing & Munro (2002), Burgess & Spencer (2000), and MacDonald (2002) conducted in Canada, Britain and Australia respectively, all show that the phenomenon noted by Otto Jespersen in 1904 is still going strong. L2 teachers are often worried that they aren't well prepared to teach pronunciation, or even to incorporate some pronunciation activities into their regular language classrooms. They feel as though they don't know where to start.

---

This is where research comes in. Practical research can help instructors to determine where to put the focus. Flege (1988) showed that most of the phonological changes that immigrants make in acquiring their L2 occur during the first year in the L2 environment. That is not to say that there aren't any changes after that, for Trofimovich and Baker (2006) demonstrated that there surely are, but the first year is when the most progress is made in the absence of pronunciation instruction. If that is the case, wouldn't it be helpful to have some longitudinal studies to know which aspects of pronunciation will likely take care of themselves over time? Such information would allow teachers to focus on more intransigent problems.

Consider the development of L2 vowels. Munro & Derwing (2008) collected speech samples six times in the first year that two groups of adult ESL learners were in Canada. They were speakers of Mandarin and Slavic languages (Russian, Ukrainian and one Serbo-Croatian). We extracted vowels from the samples and conducted identification tests with human listeners who classified the vowels as belonging to an English vowel category, or Other, and then we determined how many productions were classified as the intended vowel. After one year, the identification scores for the vowel in the word 'beat' were very high. The Mandarins' productions received a score of 97% and the Slavic language speakers had a score of 90%; in other words, the vowel in the word 'beat' was interpreted by listeners as the intended vowel most of the time. It would have been a waste of precious classroom time to work on this vowel with these learners. However, the vowel in the word 'bit' presented quite a different story. The Mandarin speakers' correct identification score went from 31% to 41% in their first year; the Slavic language speakers' scores on this vowel also improved fairly dramatically, going from 20% to 48%, but neither group was able to produce this vowel accurately even half the time. This vowel is therefore an ideal candidate for instruction. It has a high functional load, which means that it differentiates a large number of words, and learners, at least from these two language groups, aren't going to make sufficient improvement on their own. We have conducted a similar study with consonants and consonant clusters in word-initial and word-final position (Munro & Derwing, forthcoming), and again, we found that many segments and combinations of segments did not require any intervention. These are just a few examples to show that research does have something to contribute to what teachers do in the classroom.

## **2. Focus on Pronunciation in Teacher Education**

My second utopian goal is an increased focus on pronunciation in teacher education. Things have improved since I first taught pronunciation, as Gilbert (2005), Celce-Murcia, Brinton and Goodwin (1996) and others have published very useful resources for teachers. There are also good student resources now available for use such as Hahn & Dickerson (1999), Grant (2001), Reed and Michaud (2006) and others. Although this not an exhaustive list, we still have a long way to go. First, there is a definite need for more courses for ESL teachers. In Canada, for instance, there are very few TESL programs that offer a full course in teaching pronunciation. Not only are there not enough courses in pronunciation pedagogy, there are TESL programs that

have no requirement for even an introductory course in Linguistics, which is surprising indeed. In addition to initial training for all ESL teachers, there is also a role for ongoing professional development. Our colleagues Lynda Yates and Beth Zielinski in Australia are developing a web resource to do just that, and I hope that other universities will emulate this initiative. I've already mentioned some studies that indicate that many ESL teachers don't feel comfortable dealing with pronunciation, but what about teachers in content classes? An increased focus on pronunciation should extend to K-12 classroom teachers as well, even though they are not necessarily designated as language teachers per se. Consider this advice from a teacher publication to social studies teachers for helping their immigrant L2 students. "Adjust speech rate and enunciation. While English is a stress-timed language, many other languages, including Spanish, are syllable-timed languages. English tends to stress one or two syllables and slur the rest of the word or sentence. This means that English sounds are often unclear to some speakers of other languages. *Thus, pronouncing equally stressed words or sentences may increase students' comprehension along with adjusted speech rate*" (Cho & Reich, 2008, p. 239) (italics added). This is a case where a little knowledge can be a bad thing. Regardless of what you think about the stress-timed vs. syllable timed debate, the kind of language advocated by Cho and Reich is not what students from a diverse set of language backgrounds need to hear. Teachers should be modelling accurate pronunciation, not trying to imitate their L2 students by putting equal stress on every syllable. We want our students to be able to communicate with other people in the community; in an immigrant setting that means adopting the local version of English. They need to hear which syllables have reduced vowels and which do not. All teachers would benefit from an increased understanding of L2 pronunciation.

### **3. Appropriate Curriculum Choices**

The next goal has to do with appropriate curriculum choices. Sometimes stand-alone pronunciation classes can be helpful, particularly if there are large numbers of students who share similar difficulties – many people from different L1s will have problems with rhythm for example. However, programs may not have sufficient numbers to run stand-alone classes, or students may need to be working on other things, as well, and so need pronunciation to be integrated into general listening and speaking classes. Students should be exposed to multiple voices from a range of ages and dialects. Levis and Grant (2003) point to the lack of systematicity in the inclusion of pronunciation in general ESL classes and provide some suggestions for ways to incorporate pronunciation. It should naturally be a part of a speaking and listening class, and yet it often isn't.

### **4. Assessment**

In the USA, there are assessments for international teaching assistants, but in other types of language programs, and certainly in Canada, people tend to avoid assessing pronunciation. However, as with so many other areas of language proficiency, if pronunciation were tested, it would be taught. I understand that there may be a concern about what assessments could be used

---

for, but I doubt that any problems assessment may cause would be any worse than what happens to L2 speakers in the real world without assessment. I am in favour of the development of assessment tools for pronunciation. We know there will be washback if there are tests, so it is important to design good ones.

### **5. A Focus on Intelligibility and Comprehensibility**

The next goal is a focus on intelligibility and comprehensibility, rather than accentedness. These terms have distinct meanings. Accentedness is a judgment of how much one's speech differs phonologically from the local variety. It is often measured on a Likert-like scale (e.g., 1 = no accent; 9 = extremely heavy accent). Comprehensibility is a judgment of how easy or difficult an individual's pronunciation is to understand, and it can also be measured on a scale (e.g., 1 = very easy to understand; 9 = extremely difficult to understand). Intelligibility is the degree to which a listener understands a speaker; this can be measured in several different ways, including transcriptions, comprehension questions, and listener summaries of productions. In other words, accent is difference, comprehensibility is effort, and intelligibility is actual understanding.

In his special issue of *TESOL Quarterly*, Levis (2005) described two approaches to teaching pronunciation: one of these follows the nativist principle, which holds that L2 speakers should try as best they can to replicate a native-like accent. The intelligibility principle is the basis of the other approach, which holds that L2 speakers should be comfortably understandable. Recently, Abrahamsson and Hyltenstam published an extensive study in *Language Learning* (2009), in which they examined the "nativeness" of nearly 200 individuals who spoke Swedish as a second language. Spanish was the L1 of all the participants, who started learning Swedish between the ages of 1 and 47 years. These people were selected because they self-identified as being native-like in Swedish and all were extremely high proficiency. However, when compared to native speakers in a battery of tests, none of the late learners (over the age of 11 when they started speaking Swedish) had equivalent scores. Thus, despite their exceptional language skills, these high proficiency L2 speakers were still not comparable to native speakers.

Abrahamsson and Hyltenstam's finding suggests that native speaker performance should not be the goal of L2 learners, but rather to aspire to the status of highly intelligible, easily comprehensible bilingual speakers. However, as Levis (2007) noted, many ESL teachers are still clinging to the nativist principle. Over the years, I have given many talks about intelligibility, comprehensibility and accentedness, and very often someone will tell me that his/her students want help with aspects of their accent that don't interfere with intelligibility – such as the interdental fricatives, which is invariably the prime example used. They say that their students are compromised socially because they can't make these sounds. I seriously doubt that the TH sounds are the only problem those individuals have. Many L2 users of English who don't make those sounds, and who have no other pronunciation difficulties are accepted for who they are: articulate, fluent speakers.



## **6. Useful User-friendly Software**

My next utopian goal would be to see more development of easy-to-use and useful software. We have some very good researchers in our field who have worked with technology, and there are some good programs, but nowhere near enough. Teachers have been encouraged to use resources such as PRAAT more than they do. This is unrealistic, however, because PRAAT was designed for researchers, not for classroom teachers. Only the most conscientious pronunciation teacher is going to tread in that territory, so it is necessary that more teacher-friendly resources be developed. The key benefit of computer assisted pronunciation training is that it can be individualized to the student's needs, but as Levis indicated (2007), by and large, this hasn't happened. Most commercial programs are still of the one-size-fits all variety, which means that inevitably, students are going to waste considerable time if they go through them in lockstep fashion. The problem here is that the market, with some notable exceptions, has focused on the look of programs – the bells and whistles, rather than the linguistic needs of the learners. Ideally, software should be developed that the teacher could easily customize to his or her students' specific needs for individual practice.

Virtual worlds, such as Second Life, could be ideal places for learners to go to get practice listening if they are a bit intimidated by real life circumstances. There are places in Second Life where people can go now, but it would be more beneficial if there were places designed specifically to expose learners to particular aspects of pronunciation that give them the most difficulty. In a utopian world, automatic speech recognition would give learners the feedback they need, but unfortunately, it is unlikely to reach an accurate enough level for some time. There are current technologies, such as Skype and iChat, that have great promise for opportunities to practice speaking comprehensibly with real listeners, but the extent to which the average ESL teacher utilizes these resources to enhance pronunciation is a question in my mind.

## **7. A Focus on the Native Listener**

In a utopian world, at least an ESL, immigrant-receiving world, we would put more emphasis on helping native speakers to understand accented English. It is ironic that we expect L2 learners from many linguistic backgrounds to understand each other, as well as a full range of English dialects, while at the same time, some native speakers make no adjustments for their L2 interlocutors.

A study of major impact in this area is that of Rubin (1992), who had two classes of psychology students listen to a mini-lecture recorded by a woman whose own dialect of American English was the same as that of the undergraduate students. However, in one class the students were shown a photo of a Chinese woman and were told that she was the lecturer, and in the other class, the listeners saw a picture of a Caucasian woman. The students in the class who saw the Chinese photo actually understood less of the lecture than the students who saw the Caucasian

---

woman, and they complained about her accent! Yet the researcher had used the same recording in both classes. This tells us that expectations can have a major impact on listeners.

There are people in this world who are biased against immigrants, biased against people of another race, biased against accents and essentially biased against difference of all types. There is probably not much one can do about those individuals, but there are also many people who are not anti-immigrant and not racist, but who are afraid to talk to L2 speakers because they don't think they have the skills to understand accented speech. In a study that my colleagues and I conducted (Derwing, Rossiter & Munro, 2002), we trained social work students to listen more carefully to L2 speech. The participants changed their attitudes towards their own ability to understand accents. By the end of a term, several reported a willingness to interact with L2 speakers that they hadn't felt before, and they also indicated having experienced success in real life encounters, which they attributed to the training.

In a more recent study, Kang, Rubin and Pickering (2008) designed an intervention in which university students met with international teaching assistants to do a puzzle together and share a pizza. Members of this group were more empathetic towards the teaching assistants after this intervention.

So what can native speakers do to improve communication with a nonnative speaker? Not only can they make more efforts to listen to accented speech, but they can follow the suggestions from the Københavns Sprogcenter, Dansk for Udlændinge (Danish for foreigners). The Sprogcenter produced posters that are distributed to workplaces all over Denmark, with the following 10 tips for encouraging successful interaction of native speakers with nonnative speakers:

1. Imagine what it's like to be in your colleague's shoes.
2. Involve your colleague in conversation.
3. Take time to listen.
4. Look at the person you're speaking to.
5. Say it in a different way if you are not understood.
6. Helping find the missing words.
7. Speak straightforwardly (e.g., don't use much slang).
8. Speak in a suitable tempo.
9. Don't mumble.
10. Give only a few instructions at a time.

## **8. No More Scapegoating of Accent**

The next goal is no more scapegoating of accent. A couple of years ago, the University Teaching Services (UTS) unit at my university had several distressed international teaching assistants and professors who were devastated by the very poor teaching evaluations they had

received, all of which blamed their L2 accents. These individuals came to UTS for help, and UTS approached me and two of my colleagues, one from Drama and one from Speech Pathology, to see if we could assist them.

We recruited six volunteers from the group that had approached UTS and videoed them in their own classes at the beginning of term, then worked with them once a week for 10 weeks. We videoed them again at the end of term, and compared their teaching evaluations pre and post intervention. The results were overwhelmingly positive (Derwing, Moulton, Campbell & Dumas, forthcoming). What is important, though, is that we did very little work on the participants' pronunciation. When we analysed the initial videos, we determined that most of the speakers were actually quite comprehensible from the outset. With one exception (an individual who had a persistent stutter in both L1 and L2) their problems had more to do with limited teaching skills. We focused on presentation skills (such as making eye contact with the students), and pedagogical skills (such as using a handout instead of the blackboard). At the end of the term, one of our participants received a teaching award; the change in his performance was amazing. We regret that we didn't apply for ethical clearance to show the videos publically, because overall, the before and after differences were so dramatic. But there was very little change in pronunciation, because we didn't work on that, other than focusing on projection, and ensuring that the participants knew the appropriate word stress patterns for the key vocabulary in their fields. Our participants' own undergraduate students had blamed L2 accent in their course evaluations, because accent was so salient, and it blinded them to what the real problems were.

There are two other aspects of language that contribute to a lay listener's sense that an L2 speaker has a difficult-to-understand accent. One of these is pragmatics, or knowing what is appropriate to say in a particular context. If someone uses unexpected phrases or lexical items, they may not be understood, because of the generally high predictability of much of our everyday language. The listener expects to hear one thing, and when something else comes along, he or she can't understand it. Unexpected grammatical patterns, too, can cause problems that will be blamed on pronunciation (Varonis & Gass, 1982). Work on appropriate use of language (e.g., Yates, 2004) may well result in a perception of improvement in pronunciation.

## **9. Better Strategies for Integrating Newcomers into the Community**

Some may question what integration strategies have to do with pronunciation teaching, but the extent and quality of exposure that speakers have to their L2 affects their comprehensibility (Derwing, Munro & Thomson, 2008). So how can we help newcomers increase their opportunities for speaking? Certainly in the ESL programs in Canada, we could concentrate more on conversational strategies while people are in their language classes. The focus right now is heavily weighted to grammar, reading and writing, but if people came out of those courses with stronger speaking skills, they may have a heightened willingness to communicate. ESL courses could have built-in supports, such as ethical volunteering opportunities, as recommended by Dudley (2007), which would benefit L2 learners and cooperating institutions

---

alike. The workplace, too, could become more L2-friendly and forward-looking companies already realize that it is in their best interests to encourage good communication among employees. If we consider the Mandarin and Slavic language speakers in our longitudinal study (Derwing, Munro & Thomson, 2008), a major difference between the two groups was the amount of exposure they had to English on a daily basis, including conversations at work and with neighbours, TV viewing, radio and movies. The Slavic language speakers showed improved comprehensibility and fluency over time, in the absence of instruction, whereas the Mandarin speakers did not improve. Clearly more interaction can enhance comprehensibility and fluency. For one thing, it provides for more opportunities for noticing, just as in other aspects of L2 acquisition.

## **CONCLUSION**

Perhaps in setting these utopian goals, I was aiming a little low. The goals I have laid out are actually less than utopian, and ultimately achievable. I think that our field is poised to make a significant and lasting difference right now. There have been more PhD students studying L2 pronunciation in the last three years than I can remember in the 15 years before that. It is conceivable that most teacher preparation programs could introduce at least some focus on L2 pronunciation issues, and that pronunciation could be better incorporated into L2 curricula, and better assessed. We are now at a point where most L2 teachers recognize that there is nothing wrong with having an accent, and that intelligibility and comprehensibility should be the goals of L2 speakers, not native-like status. Technology is advancing; there is a real role for virtual worlds and other sorts of practice opportunities, informed by research. Finally, there are always at least two people involved in real communication and both sides should be striving to achieve communicative success, rather than putting all the responsibility on the shoulders of the L2 speaker. Here I am addressing an ESL reality, as opposed to English as an international language. Native speakers need to loosen up a little, and make a bit more effort. We in the field of pronunciation teaching and research are the people best equipped to help them. For those of us who teach at universities and colleges, we can start with our own students. We are on the verge of a major shift in attitudes and it is our job to speed up change. In conclusion, perhaps this talk was wrongly titled after all. Instead of “Utopian Goals for Pronunciation Teaching,” perhaps it should simply have been called “Our To-Do List.”

## **ACKNOWLEDGEMENTS**

Much of the research discussed here was conducted with my colleague and friend, Murray Munro. I would also like to acknowledge Ron Thomson and Lori Diepenbroek, who worked on some of these studies. Thanks also to Lynda Yates, John Levis and Bruce Derwing for their comments. I am grateful to the participants in our studies. Some of the research reported here

was supported by the Social Sciences and Humanities Research Council of Canada and the University of Alberta.

### **ABOUT THE AUTHOR**

Tracey Derwing is a professor of TESL in the Department of Educational Psychology at the University of Alberta, and Co-Director of the Prairie Metropolis Centre for Research on Immigration, Integration and Diversity. In collaboration with Murray Munro she has carried out extensive research on second language learners' oral fluency and pronunciation, particularly the extent to which accent interferes with intelligibility. In addition, she has studied the effects of modifications native speakers make when addressing low proficiency speakers of English. She has also worked with colleagues on large studies of the settlement experiences of refugees in Canada. Most recently, she has conducted studies on English in the workplace, from the perspectives of both immigrant and Canadian-born employees. Email: [tracey.derwing@ualberta.ca](mailto:tracey.derwing@ualberta.ca)

### **REFERENCES**

- Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59, 249-306.
- Breitkreutz, J., Derwing, T.M., & Rossiter, M.J. (2002). Pronunciation teaching practices in Canada. *TESL Canada Journal*, 19, 51 – 61.
- Brown, A. (1991). Functional load and the teaching of pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: A book of readings*. London: Routledge.
- Burgess, J. & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28, 191-215.
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.
- Cho, S. & Reich, G.A. (2008) New immigrants, new challenges: High school social studies teachers and English language learner instruction. *Social Studies*, Nov/Dec., p. 239.
- Deng, J., Holtby, A., Howden-Weaver, L., Nessim, L., Nicholas, B., Nickle, K., Pannekoek, C., Stephan, S., & Sun, M. (2009). English pronunciation research: The neglected orphan of
-

- second language acquisition studies. *Prairie Metropolis Centre Working Paper Series*, WP05-09, Edmonton, AB.
- Derwing, T. M., Moulton, B., Campbell, M. & Dumas, J. (in preparation) Accent as scapegoat: Improving the instruction of international teaching assistants and professors.
- Derwing, T. M. & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39, 379-397.
- Derwing, T. M., Munro, M. & Thomson, R. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359-380.
- Derwing, T. M., Rossiter, M. J. & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development*, 23, 245-259.
- Dudley, L. (2007). Integrating volunteering into the adult immigrant second language experience. *Canadian Modern Language Review*, 63, 539-561.
- Flege, J. E., (1988). Factors affecting the degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 91, 370-389.
- Gattegno, C. (1976). *The common sense of teaching foreign languages*. New York: Educational Solutions.
- Gilbert, J. (2005). *Clear speech: Pronunciation and listening comprehension in American English*, (3rd Edition). New York: Cambridge University Press.
- Graham, C. (1978). *Jazz chants*. New York: Oxford University Press.
- Grant, L. (2001). *Well said: Pronunciation for clear communication*. Boston, MA: Heinle & Heinle.
- Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201- 223.
- Hahn, L. D. & Dickerson, W. B. (1999). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor: University of Michigan Press.
- Hultzén, L. S. (1965). Consonant clusters in English. *American Speech*, 40, 5-19.
- Jespersen, O. (1904). *How to teach a foreign language*. London: George Allen & Unwin, Ltd.
- Kang, O., Rubin, D. & Pickering, L. (2008, March). Impact of rater characteristics and prosodic measures of accentedness on ratings of international teaching assistants' oral

- performance. Paper presented at the meeting of the American Association of Applied Linguistics, Washington DC.
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184-202.
- Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39, 369-377.
- Levis, J. & Grant, L. (2003). Integrating pronunciation in to ESL/EFL classrooms. *TESOL Journal*, 12, 13-19.
- MacDonald, S. (2002). Pronunciation—views and practices of reluctant teachers. *Prospect*, 17 (3): 3–18.
- Munro, M. J. & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58, 479-502.
- Munro, M. J. & Derwing, T. M. (forthcoming). Implications of naturalistic segment development for pronunciation teaching technology.
- Nilsen, D. L. F. & Nilsen, A. (1971). *Pronunciation contrasts in English*. New York: Simon & Schuster.
- Price, O. (1665). *The vocal organ*. Menston, Yorkshire: Scholar Press, A Scholar Press Facsimile.
- Reed, M. & Michaud, C. (2006). *Sound concepts: An integrated pronunciation course*. New York: McGraw-Hill.
- Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33, 511-31.
- Stevick, E. (1957). *Helping people learn English: A manual for teachers of English as a second language*. Nashville TE: Abington Press.
- Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1-30.
- Varonis, E. & Gass, S. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition*, 4, 114-136.
- Walker, J. (1787). *The melody of speaking delineated*. London: Printed for the Author and sold by G.G.J. & J. Robinson in Pater Noster Row & T. Cadell in the Strand.
-

Yates, L. (2004). The 'secret rules of language': Tackling pragmatics in the classroom. *Prospect*, 19, 3-21.



# **The Effects of Phonetics Training on the Intelligibility and Comprehensibility of Native Spanish Speech by Second Language Learners**

**Jennifer Rasmussen**

**Mary L. Zampini**

Foreign Language  
Le Moyne College

Traditionally, second language (L2) phonetics training has been used primarily as an aid to pronunciation improvement for L2 learners. The impact that such training has on L2 listening comprehension, on the other hand, has not received systematic attention in the literature. This paper addresses this issue by presenting a study that examined the impact of phonetics training on the intelligibility and comprehensibility of native Spanish speech as perceived by L2 learners. Two learner groups (control, experimental) participated. For the pre-test, both groups listened to sentences produced by native Spanish speakers and wrote down what they said (a measure of *intelligibility*). In addition, they rated how easy they thought the speaker was to understand (a measure of *comprehensibility*). The experimental group then received six weeks of instruction on specific phonetic characteristics of the Spanish dialect spoken by the native speakers and engaged in focused listening and pronunciation practice. Both groups then took the post-test, which was identical to the pre-test. The results demonstrated that the experimental group showed significant improvement with respect to the intelligibility of some of the phonetic aspects trained. This research thus illustrates the benefits of phonetics training for helping L2 learners improve listening comprehension skills.

## **INTRODUCTION**

Listening comprehension is a central component of second language (L2) acquisition, yet, as Omaggio Hadley (2001, p. 184) notes, research on the development of L2 listening comprehension skills is “still in its infancy” (see also Rubin, 1994, for an overview of research in this area). The listening process involves numerous skills (Richards, 1983; Omaggio Hadley, 2001), one of which is the ability to discriminate individual speech sounds. An understanding of the L2 sound system is therefore critical for the development of L2 listening comprehension. Indeed, Arteaga (2000) argues that phonetics teaching in the L2 classroom is “an essential ingredient in improving students’ comprehension” (p. 342). The impact that phonetics training has on L2 listening comprehension skills, however, has not received systematic attention in the literature. In addition, and perhaps because of this, textbooks and instructional techniques largely

---

ignore the relationship between phonetics and L2 listening comprehension. For example, Arteaga (2000) examined the phonetics content of ten first-year Spanish textbooks and found that the trend was to provide “minimal coverage” of phonetics and pronunciation. In addition, an examination of more advanced textbooks in L2 Spanish phonetics reveals that the written content and practice exercises focus primarily on the *articulation* of L2 speech sounds, with little or no practice on the *perception* (comprehension) of L2 speech (e.g., Torrejón, 2000; Schwegler & Kempff, 2007; Piñeros, 2008). Thus, L2 phonetics training, when used, is typically employed as an aid to pronunciation improvement, or foreign accent reduction, rather than the development of listening comprehension skills.

Previous research has nevertheless shown that listening comprehension and L2 pronunciation are related (e.g., Postovsky, 1974; Oyama, 1982). In addition, several researchers have examined the kinds of training techniques that are most effective in helping L2 learners perceive or produce particular L2 contrasts accurately (see Bradlow, 2008, for an overview). Much of this research, however, has focused on L2 listeners’ ability to perceive a particular contrast in individual L2 words, rather than global listening comprehension of native L2 speech. There has also been a great deal of research on the intelligibility and comprehensibility of spoken speech, but most of it has focused on L2 speech as perceived by L1 listeners, rather than native speaker speech as perceived by L2 learners (e.g., Munro & Derwing, 1995; Derwing & Munro, 1997; see Munro, 2008, for an overview).

In one related area, Bent and Bradlow (2003) found that L2 learners comprehend the L2 speech of other learners of the same L1 more easily than that of a native speaker of the L2, a phenomenon known as the interlanguage speech intelligibility benefit (ISIB). Moreover, they also found evidence for a “mismatched interlanguage speech intelligibility benefit,” whereby speakers from different L1 backgrounds who speak the same L2 also found each other more intelligible than native speakers of the L2 (but see Hongyan & van Heuven, 2007, for different results).

## RESEARCH QUESTIONS

This paper contributes to the existing literature and attempts to fill a gap by examining the impact of phonetics training on the intelligibility and comprehensibility of native Andalusian Spanish speech as perceived by L2 learners. Andalusian Spanish is spoken in southern Spain, in the region of Andalusia, of which Seville is the capital city. For the purposes of this study, we employ the definition of intelligibility given by Smith and Rafiqzad (1979), who define it as the “capacity for understanding a word or words when spoken/read in the context of a sentence being spoken/read at natural speed” (p. 371). Second, we use Derwing and Munro’s (1997) definition of comprehensibility, which they consider to be the listener’s “*perception of intelligibility*” (p. 2). Given previous research findings, the following hypotheses guided this study:

1. L2 listeners with phonetics training will show greater improvement in the intelligibility of native Spanish speech than L2 listeners without training.
2. L2 listeners with phonetics training will show greater improvement in the comprehensibility of native Spanish speech than L2 listeners without training.

## **RESEARCH DESIGN**

### **Subjects**

The subjects were two groups of native English-speaking learners enrolled in third and fourth year Spanish courses at Le Moyne College. Because of the small student population at the College, we were unable to restrict our subject pool to one course. Thus, the subjects were enrolled in either SPN-201 (Intermediate Conversation and Composition), which is a third year bridge course between the lower and upper levels of instruction, or SPN-301 (Advanced Conversation). Twenty students initially participated in the study, and they were assigned randomly to one of two groups: experimental or control. The groups were balanced with respect to course enrollment and gender; however, four of the control group subjects failed to complete the post-test and therefore had to be eliminated from the analysis. None of the subjects had studied abroad, nor spent extended periods of time in a foreign country. In addition, none of the subjects reported having had an instructor from Spain.<sup>1</sup>

### **Materials and Procedure**

The procedure involved a pre- and post-test for both L2 listener groups, with a six-week training session in between for the experimental group. The materials were the same for both the pre- and post-tests and consisted of 35 sentences--3 practice sentences and 32 test sentences--produced by four native Andalusian speakers from Seville, 2 male and 2 female. The sentences contained only common vocabulary items and grammar that most second-year students of Spanish would be expected to know. In addition, the sentences were created so as to elicit certain characteristics of the Andalusian dialect (see the Appendix A for the list of sentences used). The 32 test sentences were balanced so that the subjects heard eight different sentences produced by each of the four native speakers. In addition, the sentences were randomized so that the same speaker's voice did not appear twice in a row.

---

<sup>1</sup> Beyond this basic information provided by the subjects in a background questionnaire, we do not know the extent to which they may have been exposed to the Andalusian dialect, either inside or outside the classroom. The training was provided by the first author, who had spent a year in Seville, Spain, and thus provided the experimental group with additional exposure to characteristics of this dialect.

---

A response sheet was created, in which the first sixteen sentences of the test were in the form of fill-in-the-blank, similar to a cloze test, with blanks that targeted words featuring the phonetic aspects in question (see Appendix B for a sample portion of the response sheet). For the last sixteen sentences, no part of the sentence was given, and the listeners were asked to write out the sentences in standard orthography. In addition, the subjects were asked to rate each of the 32 sentences on a Likert scale of 1 to 7 based on how difficult the speaker was to understand (1=very difficult to understand, 7=very easy to understand).

For the pre-test, the experimental and control group subjects came together to a quiet room and were asked to listen carefully to each sentence and write down to the best of their ability either the missing words in standard orthography or the entire sentence, and then to rate the comprehensibility of each sentence according to the scale provided. The subjects were first given three practice sentences to transcribe and rate, and were then given an opportunity to ask any questions they had about the procedure. The experimenter then presented each of the 32 test sentences once and paused the recording until she saw that all subjects had finished writing. The session lasted approximately 25 minutes.

One week after the pre-test, the ten subjects in the experimental group returned to begin phonetics training. The training consisted of one thirty-minute session per week for six weeks. Most of the training occurred in Spanish, although English was used occasionally to clarify specific points. During week 1, the subjects were given an overview of Spanish articulatory phonetics and syllable division. In weeks 2- 6, they studied four salient pronunciation characteristics of the Andalusian dialect, as follows:

- Aspiration or deletion of syllable-final /s/: In many Spanish dialects, including Andalusian, an /s/ at the end of the syllable may be pronounced like an aspirated [h] or deleted altogether (e.g., *estás* ('you are'): [ehtáh] or [etá]).
- Synalepha: Spanish exhibits linking or elision of vowels across word boundaries. This study focuses only on the linking of identical vowels across word boundaries, as in *la amiga* ('the friend') [lamiya]. This feature is found in all dialects of Spanish, but is one that native speakers of English struggle to acquire and was therefore included here.
- The presence of the interdental voiceless fricative phoneme, /θ/: In many dialects of Spain, orthographic 'c' (before /e, i/) and 'z' are pronounced as an interdental voiceless fricative (e.g., *cena*, *cita*, *zapato* [dinner, date, shoe, respectively]). In most Spanish dialects of the Americas, on the other hand, there is no /θ/, and these graphemes correspond to the alveolar voiceless fricative /s/. Thus, L2 learners are unaccustomed to hearing the /θ/ in Spanish unless exposed to dialects from Spain.
- Deletion of intervocalic or word-final /d/: (e.g., *tomado* ('taken'): [tomáo], and *usted* ('you'): [uhté]).

These characteristics were presented in Weeks 2 – 5 of training based on their average scores on the pre-test. That is, the aspect on which the subjects performed most poorly, synalepha, was taught first, followed by /s/ aspiration, words containing /θ/, and finally, /d/ deletion. At the beginning of each session a short review was given of the previous session. Next, the new phonetic characteristic was introduced and explained. Subjects listened to recordings of the characteristic in isolated words or phrases, and where applicable, the Andalusian pronunciation was contrasted with standard general American Spanish pronunciation. This was followed by both listening and pronunciation practice. The listening exercises progressed from isolated phrases and sentences to longer discourse chunks, often dialogues or interviews found on YouTube. All materials used during the training sessions were different from the sentences of the pre- and post-tests. Week six, the final week of training, was a review session covering all four phonetic aspects, and the post-test was administered one week after the last phonetics training session. As mentioned, the post-test was identical to the pre-test.

## RESULTS

### Intelligibility

The results presented here include only the first 16 sentences on the pre- and post-test, in which the subjects were asked to fill in blanks. Each target word was categorized by the target phonetic features it contained, and the subjects were given a point each time they correctly transcribed the word for the target feature. A Chi-square analysis was then performed for each phonetic feature to see if significant changes emerged over time for each subject group.<sup>2</sup>

**Synalepha.** As shown in Table 1 and Figure 1, the control group transcribed 3% ( $n=1$ ) of synalepha occurrences correctly during the pre-test and 11% ( $n=4$ ) correctly in the post-test. This difference is not statistically significant ( $\chi^2_{(1)} = .166, p > .05$ ). The experimental group did not transcribe any instance of synalepha correctly during the pre-test and 20% ( $n=12$ ) correctly during the post-test. This difference proved statistically significant ( $\chi^2_{(1)} = 10.191, p < .01$ ). Thus, these results confirm the first hypothesis, which predicted that the experimental group would show greater improvements in intelligibility than the control group as a result of training.

---

<sup>2</sup>Intervocalic /d/ deletion was eliminated from the analysis, because it was not produced enough by the native speakers to provide useful data. /d/ deletion is variable and appears more frequently in informal speech. For the present experiment, however, the native speakers were asked to read sentences, and despite attempts to get them to produce them as naturally as possible, they did not produce all target features in every context in which they could potentially appear. The remaining discussion, therefore, does not include this feature.

---

Table 1. Pre- and post-test synalepha scores.

	Total Possible	Pre-test Scores	Percent Correct	Post-test Scores	Percent Correct	p-value
<b>Control</b>	36	1	3%	4	11%	.166
<b>Experimental</b>	60	0	0%	12	20%	.001*

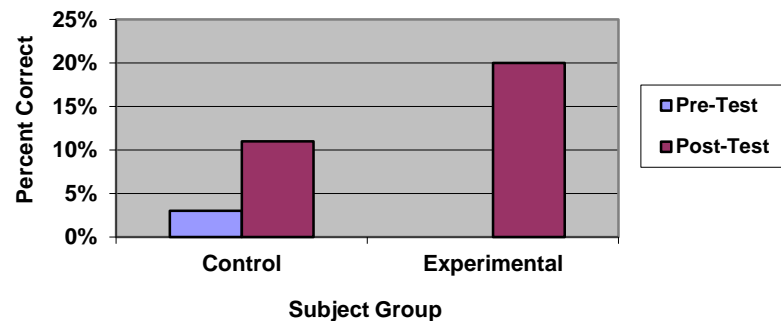
\* $p < .05$ 

Figure 1: Pre- and post-test synalepha scores for control and experimental groups

*/s/ aspiration.* As shown in Table 2 and Figure 2, the control group transcribed 15% of */s/* aspiration correctly during the pre-test and 25% correctly in the post-test. This difference is statistically significant ( $\chi^2_{(1)} = .449, p < .05$ ). The experimental group transcribed 14% of */s/* aspiration correctly during the pre-test and 41% correctly during the post-test. This difference is also statistically significant ( $\chi^2_{(1)} = 30.749, p < .001$ ). Thus, both groups showed improvement in the comprehension of words containing aspirated */s/*, and therefore, seem to contradict the first hypothesis. The gains made by the experimental group do appear to show a trend toward greater improvement than the control group, however. A between-group analysis was therefore conducted to see if the observed differences between the two groups proved significant, but they did not, so the first hypothesis could not be confirmed on the basis of this data.

Table 2. Pre- and post-test /s/ aspiration scores.

	Total Possible	Pre-test Scores	% Correct	Post-test Scores	% Correct	p-value
Control	102	15	15%	26	25%	.015*
Experimental	170	23	14%	70	41%	.000*

\*p&lt;.05

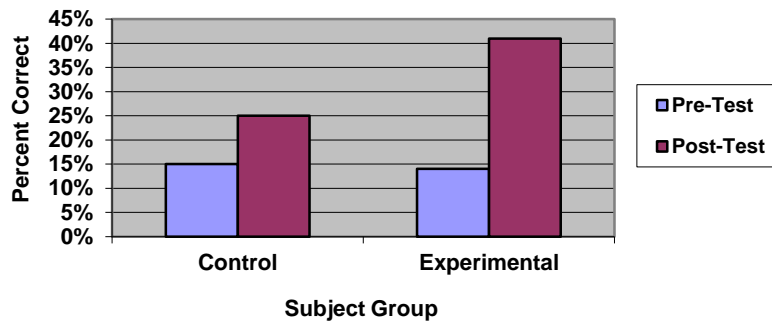


Figure 2. Pre- and post-test /s/ aspiration scores for control and experimental groups.

**Words containing /θ/.** As shown in Table 3 and Figure 3, the control group transcribed 44% of words containing /θ/ correctly during the pre-test and 39% correctly in the post-test. This difference is not statistically significant ( $\chi^2_{(1)} = .608, p > .05$ ). The experimental group transcribed 47% of the words containing /θ/ correctly during the pre-test and 57% correctly during the post-test. This difference was not statistically significant ( $\chi^2_{(1)} = .666, p > .05$ ) for the experimental group either, nor did the between-group analysis prove significant. Thus, these results also fail to confirm the first hypothesis. There appears to be a slight upward, positive, trend on the part of the experimental subjects, but no strong conclusions can be made on the basis of these results.<sup>3</sup>

<sup>3</sup> An anonymous reviewer wondered whether the trend indicating improvement in the comprehension of words containing /θ/ may have been due to the fact that this sound also appears in English. This may indeed be a contributing factor for the experimental group, who had been explicitly taught that the Spanish /θ/ is comparable to English, albeit with a different orthographic representation. It is important to note, however, that the mere presence of /θ/ in English does not appear to facilitate acquisition of this sound in L2 Spanish. After all, the control subjects did not show improvement, and overall comprehension by both groups is still somewhat low. The differing

Table 3. Pre- and post-test scores for words containing /θ/

	Total Possible	Pre-test Scores	% Correct	Post-test Scores	% Correct	p-value
Control	18	8	44%	7	39%	.435
Experimental	30	14	47%	17	57%	.414

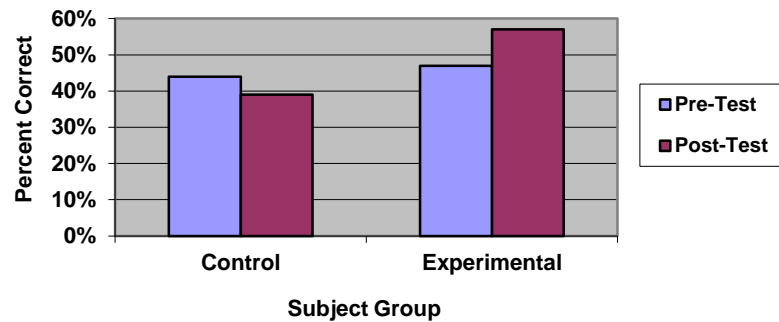


Figure 3. Pre- and post-test /θ/ scores for control and experimental groups.

### Comprehensibility

Recall that the comprehensibility ratings were based on a Likert scale of 1 to 7, where 1 was “very difficult to understand” and 7 was “very easy to understand.” The subjects’ ratings were analyzed for the interaction between the pre- and post-test vs. the control and experimental groups in a 2x2 Mixed Design Factorial Analysis of Variance (ANOVA). As shown in Table 4, the mean comprehensibility rating for the control group in the pre-test was 3.63 ( $SD=.44$ ) and in the post-test was 4.03 ( $SD=.62$ ). The mean rating for the experimental group in the pre-test was 3.23 ( $SD=.83$ ) and in the post-test was 4.03 ( $SD=.58$ ). The ANOVA thus showed no significant interaction between the pre and post-test and the control and experimental group. Moreover, the mean differences between the control and experimental groups were not significant. However, there was a significant difference in the average comprehensibility ratings between the pre-test and the post-test ( $F(1,14)=13.60, p<.01$ ) when all subjects were grouped together. Overall, the mean comprehensibility rating for the pre-test was 3.38 ( $SD=.72$ ) and for the post-test was 4.03 ( $SD=.57$ ). In other words, although there was no significant difference between the control and

orthographic representations for /θ/ in the two languages seems to have an inhibitory effect for many English-speaking learners of Spanish.



experimental groups, both groups gave significantly higher comprehensibility ratings during the post-test than in the pre-test. These results, therefore, fail to support the second hypothesis, which predicted that the experimental group would show greater improvements over time in their comprehensibility ratings than the control group.

*Table 4. Mean comprehensibility ratings based on a scale of 1 (“very difficult to understand”) – 7 (“very easy to understand”)*

	Pre-test	Post-test
Control	3.63	4.03
Experimental	3.23	4.03
<b>Both groups combined</b>	<b>3.38</b>	<b>4.03</b>

## DISCUSSION

To summarize, the results of this study reveal some support for the hypothesis that phonetics training improves the intelligibility of native Spanish speech for L2 learners. The synalepha results provide clear support for the hypothesis, in that the experimental group showed significantly better comprehension of sequences containing synalepha in the post-test than in the pre-test, whereas the control group did not. The results for /s/ aspiration and words containing /θ/ failed to support the first hypothesis, since both subject groups performed significantly better in the post-test with regard to /s/ aspiration, but neither group showed significant improvements with regard to words containing /θ/. In both cases, however, the results suggest a positive trend by the experimental group vis-à-vis the control group, but additional testing with larger subject groups would be needed to confirm this trend.

The results for /s/ aspiration were unexpected, in that the control group showed improved comprehension during the post-test. A surface overview of the transcription of the target words, however, indicates that the control group may have accurately transcribed more function words (such as the definite articles *los, las*) in the post-test than the pre-test. Thus, an interesting follow-up would be to see whether or not significant differences emerge with respect to content words, as opposed to function words. In addition, since all subjects were enrolled in relatively advanced Spanish classes throughout the course of the experiment, it is possible that they all received some exposure to speakers that exhibit /s/ aspiration in their interactions with instructors, classmates, and other native speakers.

With regard to words containing /θ/, the lack of significantly improved comprehension over time by the experimental group may have been due to the fact that this was one of the last features taught and practiced; as a result, the subjects had less time to assimilate this characteristic. It is also interesting to note that the /θ/ was not produced consistently by the native speakers who read the sentences used in the tests; thus, there were fewer /θ/ tokens in the speech sample than expected. Many Andalusian dialects do not distinguish between /θ/ and /s/, unlike northern Spain; they have either just /s/ (the standard Andalusian variety) or just /θ/. However, because Seville is the capital of Andalusia and a larger city with greater contact with Madrid (where /θ~/s/ distinction is the norm), many speakers from Seville also make the same distinction. Our informants reveal, however, that the distinction may not be a categorical phonemic one, but rather an instance of allophonic free variation, whereby either /θ/ or /s/ may appear in words that contain orthographic 'c' or 'z'.

As for the comprehensibility of native Spanish speech, the results here fail to confirm the hypothesis that the experimental group would show significantly higher comprehensibility ratings than the control group as a result of training. In fact, when grouped together, both groups as a whole gave significantly higher ratings in the post-test. As already mentioned, both subject groups were enrolled in Spanish courses throughout the time of the experiment; thus, it may be that all subjects simply became more comfortable over time with listening to Spanish speech as a result of their classes and coursework. It may also be that six half-hour training sessions is not enough to significantly impact listeners' confidence with respect to their ability to comprehend native speech. This topic therefore merits further investigation.

The inconsistency in the findings points to two primary limitations of this study: the small number of subjects and the variability of the phonetic features examined. Future research in this area, therefore, needs to recruit a larger number of subjects. In addition, more natural, less formal, speech samples may increase the appearance of certain phonetic features. Pre-determined sentences were used instead of unscripted speech in an attempt to ensure that numerous instances of the target features appeared in the test materials. However, that did not turn out to always be the case. An alternative might be to provide native speakers with pictures and ask them to narrate a story (Munro, 2008). Pictures would contain items or actions that would likely solicit the use of a word containing a particular phonetic feature.

## CONCLUSION

To conclude, in spite of some inconsistent findings, the results of the current study do suggest that phonetics training can improve the intelligibility of native Spanish speech for English-speaking learners. Such findings have implications for the L2 classroom with regard to how listening skills are taught and lend support to those who argue that phonetics instruction should be integrated more effectively into the L2 curriculum. In addition, the knowledge gained from studies such as this one can be used to improve short-term training programs and orientations for study abroad students, as well as for organizations that send representatives abroad. Phonetics

training on the dialect to which individuals will be exposed may help them to integrate more quickly into the immersion environment and may lessen their overall anxiety. More systematic experimental research, however, is clearly needed.

### ABOUT THE AUTHORS

Ms. Jennifer Rasmussen received her B.A. in Peace and Global Studies and Spanish from Le Moyne College in May 2009. This paper presents some of her research that she completed for her undergraduate honors thesis, *The Effects of Phonetics Training on the Intelligibility and Comprehensibility of Native Spanish Speech by Second Language Learners*, written under the direction of Dr. Mary L. Zampini.

Mary L. Zampini received her Ph.D. in Hispanic Linguistics from Georgetown University and is currently an Associate Professor of Spanish at Le Moyne College in Syracuse, NY. She has published several articles on L2 Spanish and bilingual speech production and perception, and is co-editor, along with Jette G. Hansen Edwards, of *Phonology and Second Language Acquisition*, published by John Benjamins in 2008. Email: [zampinml@lemoyne.edu](mailto:zampinml@lemoyne.edu)

### REFERENCES

- Arteaga, D. L. (2000). Articulatory phonetics in the first-year Spanish classroom. *Modern Language Journal*, 84, 339-354.
- Bent, T. & Bradlow A. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, 114, 1600-1610.
- Bradlow, A. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition* (287-308), Philadelphia: John Benjamins.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1-16.
- Hongyan, W. & van Heuven, V. J. (2007). Quantifying the interlanguage speech intelligibility benefit. *Saarbrücken*.1729-1732.
- Munro, M.J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition* (183-218), Philadelphia: John Benjamins.
- Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
-

- Omaggio Hadley, A. (2001). *Teaching language in context*, 3<sup>rd</sup> edition. Boston: Heinle & Heinle.
- Oyama, S. (1982). The sensitive period and comprehension of speech. In S. Krashen, R. Scarcella, & M. Long (eds.), *Child-adult differences in second language acquisition (20-38)*, Rowley, MA, Newbury House.
- Piñeros, C. E. (2008). *Estructura de los sonidos del español*. Upper Saddle River, NJ: Pearson.
- Postovsky, V. A. (1974). Effects of delay in oral practice at the beginning of second language learning. *The Modern Language Journal*, 58, 229-239.
- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17, 210-240.
- Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78, 199-221.
- Schwegler, A. & Kempff, J. (2007). *Fonética y fonología españolas*, 3<sup>rd</sup> edition. Hoboken, NJ: John Wiley & Sons, Inc.
- Smith, L. E. & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, 13, 371-380.
- Torrejón, A. (2000). *Introducción a la fonética y fonología españolas*. Boston: McGraw Hill.

## APPENDIX A

### Sentences used for pre- and post-tests

The 35 sentences used in the pre- and post-tests appear below, with an English translation in parentheses (note: the subjects did not see the sentences in written form, nor did they see an English translation). Three sentences were used for practice and were not included in the analysis of the results. The 32 test sentences were spoken by four native speakers of Andalusian Spanish (two male, two female)—eight sentences per native speaker—and were presented to the subjects in random order.

1. Ayer compré unos zapatos en el centro. (*I bought shoes in the center yesterday.*)
  2. El sábado fui a la universidad para estudiar con algunos amigos. (*On Saturday I went to the university to study with some friends.*)
  3. Ahora mismo estoy en mi casa pero me voy pronto. (*Right now I am at home but I'm leaving soon.*)
  4. La gente de Japón es muy simpática y amable. (*People from Japan are very nice and friendly.*)
  5. Durante la noche no hay luz. (*There is no light during the night.*)
  6. No he hecho nada para la clase de sociología. (*I have not done anything for my sociology class.*)
  7. Me dijo que Carmen es mayor que Juan. (*He told me that Carmen is older than Juan.*)
  8. Han tomado mucha cerveza, ¿verdad? (*They've had a lot of beer, haven't they?*)
  9. No vimos a nadie en la ciudad. (*We didn't see anyone in the city.*)
  10. Hay que doblar a la izquierda para llegar al correo. (*One must turn left to arrive at the post office.*)
  11. Esa mujer trabaja demasiado. (*That woman works too much.*)
  12. ¿Has comprado las entradas para la función? (*Have you purchased the tickets for the event?*)
  13. Van a jugar en el parque hoy. (*They're going to play in the park today.*)
  14. Hay muchas flores en el jardín. (*There are a lot of flowers in the garden.*)
-

15. A ellos les fascina viajar en tren. (*They love traveling by train.*)
16. Creo que Miguel está en la universidad. (*I think Miguel is at the university.*)
17. Los niños lloran cuando están cansados. (*Children cry when they are tired.*)
18. No hay nadie que pueda ayudarme. (*There isn't anyone who can help me.*)
19. La amiga de Elena llegó hoy. (*Elena's friend arrived today.*)
20. Están preocupados por el examen. (*They're worried about the exam.*)
21. No tengo nada que hacer. (*I don't have anything to do.*)
22. Esteban fue a comprar más pan. (*Esteban went to buy more bread.*)
23. Estoy harto(a) de escribir y estudiar tanto. (*I'm sick of writing and studying so much.*)
24. Queremos descansar durante las vacaciones de agosto. (*We want to rest during vacation in August.*)
25. El helado es el mejor postre de todos. (*Ice cream is the best dessert of all.*)
26. Me gustan más las uvas verdes que las moradas. (*I like green grapes better than purple ones.*)
27. No hay nada como el jamón de España. (*There isn't anything like Spanish ham.*)
28. Celebran su aniversario en marzo o abril. (*They celebrate their anniversary in March or April.*)
29. Ya han preparado los dulces para la fiesta. (*They've already made the sweets for the party.*)
30. Ayer vi treinta pájaros en el parque más o menos. (*I saw about thirty birds in the park yesterday.*)
31. A la gente española le gusta almorzar tarde. (*Spaniards like to eat lunch late.*)
32. Ya ha apagado la luz. (*He's already turned off the light.*)
33. La fiesta empieza a las nueve o diez de la noche. (*The party begins at nine or ten pm.*)
34. Ella quiere cortarse el pelo para tener un estilo más a la moda. (*She wants to cut her hair to have a more fashionable style.*)
35. No puedo salir sin llevar mis gafas de sol. (*I can't go out without wearing my sunglasses.*)

APPENDIX B

Sample portion from rating instrument

Post-Test November 2008

Subject Number: \_\_\_\_\_

**Practice**

Please write to the best of your ability what the speaker said.

1. El sábado fui a \_\_\_\_\_ para estudiar con \_\_\_\_\_.
2. Hay que doblar a la izquierda para \_\_\_\_\_.
3. Ella quiere \_\_\_\_\_ para tener un estilo más a la moda.

**Test**

Please write to the best of your ability what the reader said.

1. Ayer compré \_\_\_\_\_ en el centro.
2. Ahora mismo \_\_\_\_\_ pero \_\_\_\_\_ pronto.
3. La gente de Japón \_\_\_\_\_ y amable.
4. \_\_\_\_\_ por el examen.
5. \_\_\_\_\_ en marzo o abril.
6. \_\_\_\_\_ como el jamón \_\_\_\_\_.

Based only on the speaker's voice, how easy was the reader to understand?

Very difficult to understand

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

1 2 3 4 5 6 7

Very easy to understand

Jun, H. G. & Li, J. (2010). Factors in raters' perceptions of comprehensibility and accentedness. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1<sup>st</sup> Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 53-66), Ames, IA: Iowa State University.

---

## **Factors in Raters' Perceptions of Comprehensibility and Accentedness**

**Heesung Grace Jun**

**Jinrong Li**

Iowa State University

The purpose of this study is to examine native-speaker (NS) and non-native speaker (NNS) raters' thought processes while rating L2 speech samples for their comprehensibility and accentedness and identify the factors that may cause listeners to rate speech in certain ways. Think-aloud or vocalization of the thought processes of each speech rater was used to understand what aspects of speech and pronunciation six raters noticed while rating seven ESL learners' speech for comprehensibility and accentedness. We found that there were both similarities and differences between the factors noticed while rating for accentedness and comprehensibility. In addition, the NS and NNS raters showed some major differences in the aspects mentioned during think-aloud.

### **INTRODUCTION**

The relationship between the comprehensibility and accentedness of second language (L2) speech has been shown to be complex and significant for the teaching and assessment of L2 pronunciation. Arguably, comprehensible speech is a more reasonable and realistic goal than accent-free speech. However, it might not be easy to draw a clear-cut line between the two aspects, as research has indicated that the ease of understanding of an L2 speech sample might be affected by accented features of the speech sample (e.g., Munro & Derwing, 1995).

Several studies have revealed the complex interrelatedness of ratings of comprehensibility and accentedness of L2 speech. Generally, accentedness refers to the degree to which "the pronunciation of an utterance sounds different from an expected production pattern" (Munro, Derwing, & Morton, 2006, p. 112). On the other hand, understanding of L2 speech has been further differentiated into intelligibility and comprehensibility. While intelligibility emphasizes actual understanding, which is usually assessed by transcription tasks, comprehensibility focuses on listeners' estimation of difficulty in understanding (Munro, Derwing, and Morton, 2006, p. 112). We have chosen to focus on comprehensibility rather than intelligibility in the current study because comprehensibility has been shown to more affected by factors related to accentedness.

Munro and Derwing (1995) and Derwing and Munro (1997) examine the relationship between comprehensibility, intelligibility, and accentedness. The findings of these two studies show that there is a "quasi-independent" relationship between comprehensibility and accentedness

---



(Derwing & Munro, 1997, p. 1; Munro & Derwing, 1995, p. 73). More specifically, they found that comprehensibility and accentedness were related, but that their correlations were not very strong for most of the raters. Moreover, the strength of correlation varied among different raters. Thus, the findings seemed to provide more support for the idea that some accent features do not interfere with comprehensibility.

Derwing and Munro (1997) added open-ended questions to the rating sheet to explore the factors that might have affected raters' judgment. The study identified eight common categories of contributing factors including segmental features, grammar, speech rate, prosodic features, fluency, enunciation, speaking volume, and vocabulary. The researchers concluded that the factors were weighted differently in rating for comprehensibility and accentedness. However, it is not clear whether there might be other underlying factors affecting comprehensibility and accentedness. Furthermore, it is unclear if these findings would hold true for both native-speaker (NS) raters and non-native speaker (NNS) raters. Therefore, the purpose of this study is to further examine NS and NNS raters' thought processes while rating L2 speech samples for their comprehensibility and accentedness and identify the factors that cause listeners to rate speech in certain ways.

We chose to address this issue by using think-aloud protocols. Most recent research on raters' thought processes when they are evaluating oral skills is based on either stimulated recall (e.g., Winke, 2008), stimulated recall and post-task interviews (e.g., Isaacs & Thomson, 2009), or verbal reports, also called think-aloud protocols (e.g., Brown, Iwashita, & McNamara, 2005). Think-aloud protocols are commonly used to explore people's thought processes while completing a task, but few studies besides Zielinski (2008) have used it to explore raters' thought processes while rating for comprehensibility and accentedness. Therefore, we believe that the technique might help to notice the details of the raters' internal thought processes while rating and provide evidence for why raters rated as they did.

Specifically, the study seeks to address the following three research questions:

1. What do raters notice while rating ESL learners' speech for comprehensibility and accentedness?
2. Are there differences between aspects of pronunciation that raters notice when rating for comprehensibility and when rating for accentedness?
3. Are there differences between native-speaker (NS) and non-native speaker (NNS) raters in the aspects they notice?

## **METHOD**

### **Participants**

Speech samples were collected from seven ESL learners of different proficiency levels. Table 1 below summarizes the characteristics of the speakers. Since all of them are students at the same university which requires international students to take an English placement test upon their arrival, we classified them into different proficiency levels based on their test results. Specifically, 101C is a course designed to help undergraduate students with their academic writing skills, and 101D aims to help graduate students with their skills in writing research

---

articles. When undergraduate students are waived from 101C, they can take regular first-year composition classes with native-speakers. On the other hand, when graduate students are waived from 101D, they are not required to take any extra ESL classes.

*Table 1. Characteristics of Speakers Rated in the Study*

Speaker	Gender	L1 background	English proficiency	Level of education
1	Female	Korean/English	101C waived	Sophomore
2	Female	Chinese	101D	Graduate
3	Male	Chinese	101C	Freshman
4	Female	Chinese	101C	Freshman
5	Female	Chinese/Japanese	Intensive English	Graduate
6	Male	Indian	101D waived	Graduate
7	Female	Chinese	101C waived	Junior

Among the seven participants, Speakers 1, 6, and 7 had the highest proficiency level since their English placement test results indicated that they did not need to take extra ESL writing classes at the undergraduate or graduate level. Speakers 2, 3, and 4, on the other hand, were placed into ESL writing classes for undergraduates or graduates, suggesting that their proficiency level was lower. Speaker 5 is a graduate student; however, English is her third language, and at the beginning of the study, she was still in the intensive English program which is usually for students who do not meet the university's foreign language test requirement. Thus, Speaker 5 is at the lowest proficiency level in this group. Additionally, as shown in Table 1, five of the seven speakers are native Chinese speakers.

The raters were three NS raters and three NNS raters. Table 2 summarizes their characteristics. All have had some foreign language learning and ESL teaching experience. Four were doctoral students in applied linguistics (AL), while the other two were instructors in an intensive English program.

*Table 2. Characteristics of Raters in the Study*

Rater	Gender	L1 background	Occupation
1	Male	NS of English	ESL instructor
2	Male	NS of English	Doctoral student in AL
3	Male	NS of English	ESL instructor
4	Female	Korean	Doctoral student in AL
5	Male	Russian/Ukrainian	Doctoral student in AL
6	Male	Chinese	Doctoral student in AL

None of the raters were acquainted with the speakers, and thus, they were not familiar with the speakers' voices and/or accents. The Chinese rater might have been affected by the fact that the speakers were mostly Chinese, as studies have shown that non-native listeners' listening comprehension is affected differently depending on the native language of the speakers (Gass & Varonis, 1984; Major, Fitzmaurice, Bunta, & Balasubramanian, 2002). However, since the rater is a trained ESL teacher, we believe that the possible effect of one rater's familiarity with the

speech characteristics of the majority of the speakers on the study's results could be regarded as minimal in these circumstances.

## **Materials**

Speech samples were elicited by using a short passage from Bailey and Nunan (2005) for reading-aloud and three questions for spontaneous speech on a familiar topic (traveling, hobbies, and health). See Appendix A for the passage and questions. The 9-point rating scales for comprehensibility and accentedness were adopted from Munro and Derwing (1995) and Derwing and Munro (1997) since they have been widely used in studies on listeners' perceptions of L2 comprehensibility and accentedness. See Appendices B and C for the rating scales.

## **Procedures**

Participants were recruited on a voluntary basis. Each speaker met with one of the researchers to read the passage aloud and talk about a topic of his or her own choice. No time limit was set for the spontaneous speech. The researchers recorded the speakers' performances individually using a digital voice recorder and then imported the data into Audacity, a voice recording and editing software program. In Audacity, the researchers extracted the first thirty-second segment of each response as speech samples and prepared a total of fourteen speech samples, with two from each of the seven speakers. The speech samples were then transformed into MP3 files and arranged in random order.

The rating sessions were arranged with each rater individually. For each session, a rater came to a computer lab to listen to the speech samples, using a headset to perform the think-aloud task which was recorded by a digital voice recorder. The raters were asked to listen to the speech samples and rate them for comprehensibility. They were also asked to think aloud about factors that might have affected their comprehensibility ratings at the same time. Then, the raters listened to the recordings again and gave an accentedness rating to each recording while thinking aloud about factors that might have affected their accentedness ratings. The think-aloud files were then imported to Audacity and converted into MP3 format for transcribing and coding.

We did not provide any training to the raters ahead of time because the purpose of the study is to explore raters' intuitive perception of L2 speech. Prior training might influence the raters in their analysis of the factors underlying their rating of comprehensibility and accentedness of the L2 speech samples. Since all the raters have had exposure to ESL students, we assumed that their understanding of comprehensibility and accentedness and different degrees of the two speech aspects would be comparable.

## **Analysis**

In a pilot study, the researchers listened to raters' think-aloud data and came up with a coding scheme of tentative categories of underlying factors that might have affected raters' judgment of the comprehensibility and accentedness of speech samples. Using this coding scheme, the researchers coded the new think-aloud data after transcribing the key words in the raters' comments. The think-aloud files were separated into twenty-eight segments, with each focusing

---

on either the comprehensibility or the accentedness of the fourteen speech samples. For the first two segments, the authors listened, transcribed, and coded together, while discussing ways to handle controversial coding problems and developing rules of coding. Based on the data from this first stage, the authors made modifications to the existing coding scheme by adding new categories and deleting old ones. The final scheme can be found in Table 3. The authors then listened to and transcribed the rest of the segments (twenty-six segments) individually and later coded the data together while comparing and discussing the transcripts to have 100% agreement.

## RESULTS AND DISCUSSION

Table 3 displays the final coding scheme used for coding the think-aloud data together with the total counts for each aspect. The full coding results of the raters' think-aloud data can be found in Appendix D. Twenty-four aspects of speech were identified as having been mentioned in the think-aloud data, and these aspects were categorized into four groups—segmental features, suprasegmental features, global impression, and others. Counts were obtained for each aspect, with total and separate counts for focus of rating (i.e., whether the rater was rating for accentedness or comprehensibility) and the native vs. non-native rater distinction.

***RQ1: Aspects noticed while rating for comprehensibility and accentedness.*** The first research question asked what raters noticed while rating ESL learners' speech for comprehensibility and accentedness. In very broad terms, we can say that the raters noticed all of the twenty-four aspects in the final coding scheme, but to go into more detail, the total counts from Table 3 will be considered. The first main finding is that raters frequently mentioned ease of understanding (67) while rating for comprehensibility and accentedness (90) while rating for accentedness. However, this is somewhat of an expected outcome because we can expect raters to mention and describe the speech feature for which they are currently trying to assign a rating. The more interesting finding seems to be that while raters were rating for accentedness, they mentioned comprehensibility (33), and in the same way, when raters were rating for comprehensibility, they mentioned accentedness (22). This finding could point to the possible interrelationship or “quasi-dependence” between comprehensibility and accentedness that was suggested by Derwing and Munro (1997). Also under the global impression category, identification of L1 (31) and speech rate (20) were often noted by raters as factors influencing comprehensibility and accentedness ratings.

Other general findings were obtained from the total counts. Firstly, segmentals, particularly consonants (43) and vowels (32), had very high total counts. This could be because segmentals are more salient features in speech. Another possible reason is that these features have been studied extensively, and thus raters may already have accessible language and terms to discuss errors associated with individual sounds. Raters were often able to point out specific segmental errors such as /r/ insertion after a vowel or specific vowels and diphthongs that were mispronounced. They were even able to suggest reasons for incorrect sounds such as “...the tongue is too far forward for s's” (Rater 1).

Of the suprasegmental features, linking was the most commonly mentioned aspect (31). In contrast to when they were commenting on segmental features, when raters were commenting on suprasegmental features, their discussion tended to be more general. In most cases, raters related their general judgments from intuition rather than pointing out specific reasons for their decision

as they had done while talking about segmental features. For example, a typical comment on intonation was "...the intonation doesn't sound right..." (Rater 4). This may point to the raters' lack of precise language to discuss suprasegmental features.

*Table 3. Final Coding Scheme for Think-Aloud Data with Totals for Each Aspect*

Categories	Aspects	Total
Segmental	Consonants	<b>43</b>
	Vowels	<b>32</b>
	Syllables (schwa insertion)	4
Suprasegmental	Intonation	23
	Linking	<b>31</b>
	Stress	12
	Rhythm	15
Global impression	Enunciation	8
	"Word pronunciation"	9
	"Pronunciation"	16
	Identifying L1	<b>31</b>
	Ease of understanding (C)	67
	Comprehensibility (A)	<b>33</b>
	Accentedness (C)	<b>22</b>
	Accentedness (A)	90
	Speech rate	<b>20</b>
	Pauses/fillers	12
	Good sense of language	3
	Fluency	4
	Others	Grammar
Speech impediment/lisp		11
Type of speech (read vs. spon.)		6
Quality of recording		6
Recognition of words		12

An interesting finding from the Others category was related to grammar features. When Raters 1 and 3 rated the speech of Speaker 1, they noticed grammatical mistakes. Speaker 1 speaks very fluently, almost native-speaker-like, and this may be why the raters could shift their attention away from pronunciation and focus on grammatical aspects. This finding suggests that with high comprehensibility and low accentedness, the rater's attention can shift to grammar and the actual content of the speech samples. However, when we considered the average ratings together with the types of think-aloud comments for each of the speech samples, no general patterns could be found regarding the relationship between the ratings and the aspects noticed for each speech

sample. We did not find that students with lower comprehensibility and/or greater accentedness ratings had certain types of think-aloud comments.

**RQ2: Differences in aspects noticed for comprehensibility rating and accentedness rating.** The second research question asked whether there were differences between what is noticed when raters are rating for comprehensibility and when rating for accentedness. Table 4 shows the top aspects noticed under the two rating conditions. A few major overlaps can be found from comparisons to think-aloud results.

*Table 4. Most Noticed Features for Comprehensibility and Accentedness Ratings*

Comprehensibility rating	Accentedness rating
Consonants (21)	Consonants (22)
Pronunciation (14)	Identification of L1 (22)
Vowels (12)	Vowels (20)
Linking (12)	Linking (19)
Speech rate (12)	Intonation (16)
Recognition of words (10)	

First, consonants, vowels, and linking seemed to be noticed for both comprehensibility and accentedness. These were the overlapping factors. The remaining three aspects under the comprehensibility rating condition—pronunciation, speech rate, and recognition of words—affect the ease of understanding, and the raters could have mentioned them more while rating for comprehensibility. An example of a rater’s think-aloud comment that makes concurrent reference to speech rate and comprehensibility is “Although this person talks slowly, it’s easy to understand” (Rater 1). On the other hand, when raters were listening for accentedness, they often tried to identify the L1 or the source of the accent, e.g., “Chinese speaker who learned to pronounce /r/ like Americans” (Rater 1) or “it’s a strong South Asian, Indian, Pakistan, Bangladesh, Sri Lankan accent” (Rater 2). Furthermore, the suprasegmental features of linking and intonation seemed to have been noticed in the rating of accentedness more than in the rating of comprehensibility.

**RQ3: Differences between NS and NNS raters.** The third research question asked whether there were differences between NS raters and NNS raters in the aspects that they noticed in the learner speech. Table 5 shows the major differences between NS raters and NNS raters.

*Table 5. Differences Between NS and NNS Raters*

	NS raters	NNS raters
Consonants	13	30
Vowels	6	26
Intonation	2	21
Linking	2	29
Comprehensibility (in accentedness rating)	26	7

NNS raters noticed segmental features more often than NS raters did. In addition, NNS raters were much more aware of other pronunciation features such as intonation and linking. On the other hand, NS raters commented on comprehensibility more often than NNS raters while rating for accentedness. The NS raters often made comments on ease of understanding, such as the amount of “attention” or “extra effort” needed for comprehension and “being frustrated” because of the incomprehensibility of a speech sample. It appears that accented speech may have had a larger effect on the NS raters’ ease of understanding or comprehensibility of speech samples compared to NNS raters. Another finding was that only NS raters noticed and commented on the lisp that one of the speakers had. All ten comments were about the same speaker. The one comment coded as speech impediment/lisp from a NNS rater was “stuttering” for another speaker. Overall, the NNS raters focused more on specific pronunciation features, while NS raters were more global in their assessments, focusing on the overall impression of the speech sample or paying attention to whether they understood the intended message. These findings are based on total counts for the NS and NNS groups. Despite the fact that there are only three members per group, analysis of the raw counts for individual raters showed that no one rater highly swayed the results.

## CONCLUSION

In this study, we explored aspects of pronunciation and speech that raters attend to while rating comprehensibility and accentedness of ESL learners’ speech samples. Through the analysis of raters’ think-aloud data, it was found that there were both similarities and differences between the factors noticed while rating for comprehensibility and accentedness. In addition, the NS and NNS raters showed some major differences in the aspects mentioned during think-aloud.

There are several limitations in our study that need to be addressed. One limitation concerns the rating procedure. In our study, the selection of read-aloud speech samples from the participants was not entirely random. We selected the first 30 seconds of the recordings from everyone, and raters were already familiar with the content when they gave ratings to the later speech samples. Thus, later ratings in the randomized sequence might have been affected by previous exposure to the same content or voice. The raters also had a single sheet of paper to record all the students’ scores. This might have encouraged them to rate students against each other since they could compare their scores back and forth. Furthermore, the same segments were used for the two ratings (comprehensibility and accentedness) and the ratings were done one right after the other, so raters could remember the speakers, and practice effects and/or carryover effects were probably present. Also, the number of times a rater listened to the speech samples was not controlled for, and as a result, a couple of raters listened twice to some speech samples, although most listened only once.

Other limitations lie in data analysis and coding. Aggregate counts were used in the results, so we did not take into consideration the issue of verbose vs. reticent raters and the possible bias this may have created. In addition, in the coding process, we identified features that were difficult to categorize, for example, “sounds natural” or “speaks clearly,” and these were not included in the total counts in the final coding scheme.

Directly following from the limitations of our study, we would like to make some suggestions for future studies in the same topic area. First, regarding the rating procedure, the order of rating should be counter-balanced (accentedness → comprehensibility for one half of the raters and comprehensibility → accentedness for the other half of raters). Alternatively, to increase the reliability of the ratings and think-alouds, it might be better for a future study to select different segments for the comprehensibility ratings and the accentedness ratings or conduct separate rating sessions for comprehensibility and accentedness with at least a few days between the two rating sessions to reduce the practice effect coming from memory. Also, we would suggest using different segments from different speakers if the pronunciation is going to be evaluated for global impression. To lessen the possibility of raters comparing speakers against each other, we suggest having each score be written on a separate slip of paper so that the raters would be less likely to remember the scores they had recorded for previous speakers.

In the data analysis and coding stage, the counts for each aspect could be standardized for each rater to reflect the amount of think-aloud that each rater produced. To solve the problem of think-aloud comments that are difficult to categorize, we suggest conducting follow-up interviews to clarify what the raters meant. In addition, a comparison of the think-aloud data with the actual speech samples by the researchers or a third rater might help shed light on what went unnoticed or what was not noted by the raters, particularly aspects such as suprasegmentals that raters might not have had the language to discuss during the rating and think-aloud sessions.

Other suggestions regarding study design are for researchers to make sure that they convey the meaning of the terms “comprehensibility” and “accentedness” to the raters a priori so that they can be clear about what they are going to be rating. Rater training and calibration of standards should be provided so that raters are not comparing speakers against each other but against some standard. Furthermore, the raters may need to be provided think-aloud training beforehand. Using a few speech samples for practice would help clarify what is expected and may help to control for the different number of think-aloud comments made across raters.

## **ABOUT THE AUTHORS**

Heesung Grace Jun is a PhD student in applied linguistics and technology at Iowa State University where she teaches academic writing to international graduate students. Her research interests are second language writing and language assessment. Email: [gracejun@iastate.edu](mailto:gracejun@iastate.edu)

Jinrong Li is a PhD student in the program of Applied Linguistics and Technology at Iowa State University. Her research interests are computer-mediated communication and second language acquisition, L2 academic writing, and assessment of speaking and writing. Email: [jinrong@iastate.edu](mailto:jinrong@iastate.edu)

## **REFERENCES**

Bailey, K. & Nunan, D. (2005). *Practical English language teaching: Speaking*. New York: McGraw-Hill.



- Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks* (ETS RR-05-05). Princeton, NJ: Educational Testing Service.
- Derwing, T. M. & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 20, 1-16.
- Gass, S., & Varonis, E. M. (1984). The effect of familiarity on nonnative speech. *Language Learning*, 34, 65-89.
- Isaacs, T. & Thomson, R. (2009, March). Judgments of L2 comprehensibility, accentedness, and fluency: The listeners' perspective. Paper presented at the 31<sup>st</sup> Language Testing Research Colloquium, Colorado, Denver.
- Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36(2), 173-190.
- Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Winke, P. (2008, September). Beyond the rubric: Raters' thought processes while rating speech samples. Paper presented at the Midwest Association of Language Testers conference, Iowa City, Iowa.
- Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36, 69-84.
-

## APPENDIX A

### 1. Read-aloud

Instructions: Go through the passage and see if there is any new vocabulary. You can check with me if there is any. When you are ready, read aloud into the microphone.

When a student from another country comes to study in the United States, he has to find the answers to many questions, and he has many problems to think about. Where should he live? Would it be better if he looked for a private room off campus or if he stayed in a dormitory? Should he spend all of his time just studying? Shouldn't he try to take advantage of the many social and cultural activities which are offered? At first it is not easy for him to be casual in dress, informal in manner, and confident in speech. Little by little he learns what kind of clothing is usually worn here to be casually dressed for classes. He also learns to choose the language and customs, which are appropriate for informal situations. Finally he begins to feel sure of himself. But let me tell you, my friend, this long-awaited feeling doesn't develop suddenly – does it? All of this takes will power.

(Paragraph taken from Bailey & Nunan, 2005)

### 2. Free response questions

Instructions: Choose and answer one of the following free response questions.

- 1) If you could visit any place in the world for a month, where would you go and what would you do there?
- 2) Nowadays, there are a lot of activities, hobbies, or forms of amusement and entertainment to choose from. What do you enjoy doing more than anything else?
- 3) What are some things that people can do to take care of their health?



**APPENDIX C**

2. Please rate the speech samples for their accentedness on a scale of 9 (1 = no accent and 9 = extremely strong accent). While you are rating, think aloud about the factors that influenced your rating of the speakers' accentedness.

Speech samples	1	2	3	4	5	6	7	8	9
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									

---

**APPENDIX D**
*Aspects mentioned in raters' think-alouds*

Categories	Aspects	Comp	Acct	Total	NS	NNS
Segmental	Consonants	21	22	43	13	30
	Vowels	12	20	32	6	26
	Syllables (schwa insertion)	3	1	4	2	2
Suprasegmental	Intonation	7	16	23	2	21
	Linking	12	19	31	2	29
	Stress	5	7	12	3	9
	Rhythm	6	9	15	8	7
Global impression	Enunciation	6	2	8	3	5
	"Word pronunciation"	5	4	9	6	3
	"Pronunciation"	14	2	16	8	8
	Identifying L1	9	22	31	14	17
	Ease of understanding (C)	67	0	67	38	29
	Comprehensibility (A)	0	33	33	26	7
	Accentedness (C)	22	0	22	13	9
	Accentedness (A)	0	90	90	54	36
	Speech rate	12	8	20	6	14
	Pauses/fillers	8	4	12	5	7
	Good sense of language	0	3	3	3	0
Fluency	3	1	4	1	3	
Others	Grammar	4	2	6	5	1
	Speech impediment/lisp	4	7	11	10	1
	Type of speech (read vs. spon.)	6	0	6	4	2
	Quality of recording	6	0	6	5	1
	Recognition of words	10	2	12	2	10

---

Ingels, S. (2010). The effects of self-monitoring strategy use on the pronunciation of learners of English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1<sup>st</sup> Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 67-89), Ames, IA: Iowa State University.

---

# **The Effects of Self-Monitoring Strategy Use on the Pronunciation of Learners of English**

**Sue Ingels**

University of Illinois at Urbana-Champaign

This paper describes preliminary results of a classroom-based study on the effectiveness of training advanced second language learners to monitor and correct their non-target use of English suprasegmental features. The independent variables were three levels of self-monitoring [listening only (L), listening + transcription (LT), and listening + transcription + annotation of corrections (LTA)] and rehearsal (R). The dependent variable was an accuracy score for three suprasegmental features: message unit boundaries, primary phrase stress, and intonation. After 16 weeks of pronunciation and self-monitoring instruction, 15 international graduate students at a U.S. university produced a 5-minute mini-lecture, and then used each self-monitoring strategy for a different segment of the lecture. After self-monitoring, participants orally produced the corrected version three times. Rehearsals were recorded and the first and third were analyzed. Preliminary data analysis for seven participants indicates that learner accuracy scores increased following self-monitoring; the greatest improvements occurred after one rehearsal; L was most useful for primary phrase stress accuracy; LTA was best for message unit boundaries and LT for intonation. Accuracy scores improved the most for primary phrase stress, though individual differences were found in learners' performance following use of a particular strategy.

## **INTRODUCTION**

Learning to perceive and produce second language (L2) sounds and prosody, and understanding how they function and are implemented in the L2, are important skills for adult language learners who want to maximize their L2 intelligibility and continue to improve their oral skills following the completion of classroom instruction. Improving language skills is an ongoing process and having access to strategies that allow language learners to take control of their own L2 pronunciation learning are considered essential for achieving academic and professional success in L2 contexts (Celce-Murcia, Brinton, & Goodwin, 1996; Dickerson, 1994; Morley, 1991; and others). Thus a primary motivator for this study was the desire to identify effective strategies for improving L2 pronunciation that learners could use independently, without access to specialized resources and without technical linguistic training. Additionally, self-reports from students enrolled in my ESL courses during previous semesters of pronunciation instruction indicated that strategies such as critical listening and transcription helped learners perceive previously unknown suprasegmental features in their own production. As a second language pronunciation teacher, I wanted to gather empirical evidence to determine whether self-monitoring strategies could be used successfully by adult L2 learners.

A variety of factors influence how successful adults are in learning an L2 phonology. Some L2 learners acquire sufficient pronunciation accuracy without explicit instruction (Riney & Flege, 1998). Such learners typically are learning an L2 before puberty or are gifted adult language learners. A number of factors influence L2 pronunciation acquisition, including factors outside the learners' control, such as language aptitude, phonemic coding ability, developmental readiness, and working memory (Celce-Murcia et al., 1996; Juffs & Rodriguez, 2007), as well as factors that learners have some ability to control, such as motivation and amount of L2 exposure, instruction, and use (Lightbown & Spada, 2006). However, for the learners in this study (prospective international teaching assistants, or ITAs, who take ESL pronunciation classes), motivation, daily exposure to English, and prior classroom instruction have not been sufficient for acquiring the necessary accuracy. These learners often are unable to identify when their pronunciation is not target-like and do not notice relevant L2 features in native speaker (NS) speech, though they know their speech differs from target L2 production.

For more than 30 years, language teachers and researchers have been particularly interested in identifying what sets apart “good” language learners from the less successful ones (Oxford, 1990; Rubin, 1975, 1981; Stern, 1975; Vann & Abraham, 1990). A key research goal has been to determine the characteristics and language learning behaviors of successful learners, and the conclusion is that successful language learners often use some of the same strategies to maximize their learning.

Most research has focused on strategies used by learners for improving L2 speaking and general L2 discourse skills (Oxford, 1990). Until recently, only a handful of researchers have focused on specific strategies and techniques that learners can use to improve L2 pronunciation. The most common strategies include noticing and self-correcting (Eckstein, 2007), self-monitoring (Eckstein, 2007; Vitnova & Miller, 2002), and self-evaluating (Peterson, 1997, 2000). Writers from at least 30 years ago have agreed on the need for self-monitoring for improving general L2 speaking abilities.

The good language learner monitors his own and the speech of others. That is, he is constantly attending to how well his speech is being received and whether his performance meets the standards he has learned. Part of his monitoring is a function of his active participation in the learning process. He is always processing information whether or not he is called on to perform. He can learn from his own mistakes (Rubin, 1975, p. 47).

Dickerson (1989) emphasizes the importance of helping L2 learners develop “self-critical abilities” (p. xiii) and implement systematic “covert rehearsal” (p. xvii) specifically for pronunciation improvement. However, in the past two decades limited additional work has been completed regarding how language learners implement self-monitoring strategies in learning pronunciation and indeed whether such strategies enable learners to make their pronunciation more target-like.

The definition used in this study for language learner strategies (LLS) comes from Hsiao and Oxford (2002, p. 372), who state that LLSs are “active, conscious, purposeful, and attentive,” and from Gu (2007) who describes LLSs as tools learners use to maximize results from their L2 learning efforts. In a pronunciation class that emphasizes student control over learning, students must learn to use strategies for applying pronunciation rules and then apply these strategies in a principled manner during private practice, or “covert rehearsal” (Dickerson, 1989). The process of covert rehearsal includes daily practice, talking aloud in English, monitoring performance for specific features, comparing performance with models stored in memory, making changes in production to match the models, and practicing changes out loud (Dickerson, 2000). In the current study, the use of self-monitoring and self-correcting strategies is based on descriptions provided by Dickerson (2000) and Hahn and Dickerson (1999). Two elements are added here to the range of covert rehearsal techniques: learner use of self-recordings and self-transcription (described in the next section). These latter two were added for the following reasons: a) the strategies had been used for several semesters in our ESL classes, with only anecdotal evidence of their effectiveness; b) they fit the Hsiao and Oxford and Gu definitions of LLSs, as being “active, conscious, purposeful, and attentive” and were skills or “tools” that could be taught to learners as a way to maximize learning; c) very little research existed to indicate whether these strategies were effective for helping ESL learners improve pronunciation; d) if the strategies were indeed useful, they seemed very suitable for use by students in their private study, both during and following completion of classroom instruction (i.e., self-directed study).

### **Self-Monitoring Strategies Used in This Study and Related Research**

Three combinations of self-monitoring strategies are investigated in this study, starting with the most basic strategy (listening only), then adding a second (transcription), and then a third strategy (annotation).

***Listening (L).*** When using this strategy, students listen to short segments of their recorded speech and then attempt to correct any non-target pronunciation. For purposes of this study, the researcher divided student recordings into 1- to 2-message unit segments (typically one or two utterances, or roughly 7 to 14 words), which were presented to students as audio files on a computer. Students did not transcribe their speech nor did they see transcriptions of their speech (though the researcher transcribed all speech samples used in this task in order to determine message unit boundaries prior to setting up the task). Listening (L) required learners to focus on their L2 production at segment, syllable, word, phrase, and discourse levels. A rationale for this strategy is that, when the learner is targeting specific features such as the pitch jump or drop on the syllable receiving primary phrase stress, attention may be drawn to non-target-like pronunciation, thus facilitating bottom-up processing (Izumi, 2003). Research findings indicate that use of critical listening as a strategy for pronunciation improvement results in improved intelligibility (Acton, 1984), reduced use of epenthesis and omission of sounds (Couper, 2003), and production of more accurate word stress and vowel quality (Dickerson, 1987).



***Listening + Transcription (LT).*** When using the LT combination, learners listen critically to their own voice recordings and write down exactly what was spoken, including target-like and non-target-like pronunciation, as well as pauses, restarts, fillers, and repairs. The goal is to create an accurate written record of a speech sample, prior to the evaluation phase. The process of transcribing may allow the learner to notice target forms (Schmidt, 1993) by systematically drawing attention to important L2 pronunciation cues. Research findings indicate that use of transcription may result in improved pronunciation and grammar (Mennim, 2003, 2007) and improved grammar and vocabulary (Lynch, 2007).

***Listening + Transcribing + Annotation (LTA).*** This strategy combination takes the process one step further. Learners refer to a checklist of features to evaluate and systematically review their own transcripts, identify non-target pronunciation, and mark corrections directly on the transcript. Reviewing the transcript is an evaluatory process that learners complete after speaking, when they have the processing resources available to attend to form and its relationship to intended meaning (Trofimovich and Gatbonton, 2006). The assumption in this study is that, because the speaker's message has already been produced, the learner can shift a larger portion of the focus to the correct, corresponding form. No research has been located to date on the role of annotation as a self-monitoring aid.

***Rehearsal (R).*** After performing one of the self-monitoring tasks (L, LT, or LTA) with a given speech excerpt, learners orally produced (rehearsed) each excerpt three times. During and following each rehearsal, learners monitored and evaluated their output, with the goal of identifying modifications that were needed in subsequent rehearsals in order to achieve target-like pronunciation. Research on the effectiveness of task repetition has found evidence for self-correction of pronunciation (Lynch & Maclean, 2001). Repetition and focus on form have been found to exploit familiarity with task, form, and meaning and possibly free up processing resources (Trofimovich & Gatbonton, 2006).

***Summary.*** Except for Couper (2003) and Dickerson (1987), researchers studying self-monitoring strategies have focused on global rather than specific changes in pronunciation features and have looked at use of strategy combinations (e.g., critical listening, transcription, and annotation combined), rather than isolating the effects of each separate strategy component. More research is needed to better understand the value and effectiveness of systematic use of well-defined self-monitoring strategies for improving specific L2 pronunciation features.

### **Rationale for Selection of Suprasegmental Features for This Study**

In the ESL course that is the basis of the study (ESL 504), both segmentals and suprasegmentals are taught, though the latter are emphasized due to students' most typical needs. An emphasis on suprasegmentals also is based on recent research, which has offered support for the importance of accurate use of suprasegmentals in promoting L2 intelligibility. The use of too many pauses and inaccurately placed MU boundaries (Tyler et al., 1988, cited in Hahn, 1999); absent or

incorrect primary phrase stress (Hahn, 1999, 2004); non-targetlike intonation (Pickering, 2001; Wennerstrom, 1998); and word stress errors (Benrabah, 1997; Guion, Harada, & Clark, 2004; Kawagoe, 2003) all have been shown to negatively impact intelligibility in English. All are cues used by NSs of English, both for sentence- and word-level processing and for interpreting the meaning of utterances within discourse. ITAs commonly work with undergraduate populations who have little or no prior exposure to accented English, so achieving a sufficient level of L2 intelligibility is critical for ITA academic and professional success. Research has not yet definitively identified one of these features as most important. However, a definitive answer may not exist. Instead, what is important for intelligibility most likely depends on listener characteristics, the speaking context and its communicative demands, and the frequency of the L2 speaker's errors.

Seven suprasegmentals were included in the larger study, but the preliminary data reported here relates to three features:

- a) Message unit boundaries (utterances or short phrases separated by brief pauses and/or characteristic intonational patterns)
- b) Primary phrase stress, the most prominent syllable in a message unit, usually occurring on the last content or function word in “new” information in a phrase (“old” or “given” information is not highlighted), but it is also used to signal contrasts, contradictions, comparison, choices, and other information the speaker intends to highlight (Hahn, 1999, 2004). Syllables under PPS usually are longer in duration and higher or lower in pitch, in comparison to surrounding syllables.
- c) Learner use of three categories of intonation were investigated (fall, rise to mid-range, and rise to high range, as described in Hahn & Dickerson, 1999). Falling or “final” intonation is used for signaling completion of a thought; a rise in pitch to mid-range indicates an incomplete (“non-final”) utterance; and a rise to the high end of the pitch range signals certain question types.

MUs are used in this study as the primary unit of spoken discourse. Use of MUs offers a means for choosing equivalent speech samples from each participant. MU length varies within and across speakers, but the other suprasegmental features follow phonological rules that operate within an MU. MUs typically have one prominence (primary phrase stress) and an intonational pattern following the PPS. Because each speaker is producing a different text, speaking at a different rate, and producing varying numbers of morphological units per MU, using the MU as the primary unit of discourse was the best way to standardize the speech samples selected.

## RESEARCH QUESTIONS

The purpose of this study was to evaluate the effectiveness of adult L2 learners' use of self-monitoring strategy combinations (*L*, *LT*, *LTA*, and *rehearsing corrections aloud*) to achieve

target-like production of three English suprasegmental features: *message unit boundaries*, *primary phrase stress*, and *intonation* (as defined in Hahn & Dickerson, 1999, and in Cruttenden, 1997). The following research questions were investigated:

### 1. Effect on Overall Suprasegmental Accuracy

In what ways does strategy use (L, LT, LTA, rehearsal) result in improved pronunciation accuracy? To what extent do strategy combinations have differential effects?

### 2. Effect on MUs, PPS, Intonation Accuracy

How effective are the strategy combinations for each pronunciation target?

## METHOD

### Participants

The original group of participants included 15 international graduate students enrolled in a 16-week ESL pronunciation class during Spring 2009. The instructor was the researcher. During the second class meeting, and without the researcher present, a colleague of the researcher collected consent forms from the students who agreed to participate in the study. Participant names remained anonymous to the researcher until after final grades were submitted. All students in the class received the same instruction and completed all of the experimental tasks as part of required coursework. Fifteen of the 16 students in the course participated.

Data from seven native speakers of Mandarin (4 female, 3 male) were analyzed for this paper. All were preparing to re-take the SPEAK test (Speaking Proficiency English Assessment Kit, published by Educational Testing Service), in order to fulfill oral proficiency requirements for teaching assistants. The SPEAK test is a 20-minute oral test, offered once per semester in a computer lab. Students provide timed responses to 12 questions and recordings of their responses are scored by anonymous raters. Rating is holistic, and pronunciation accuracy (intelligibility) is a significant consideration in rating. Students who receive a failing score must take an ESL course or work with a tutor prior to taking the test again. Of the seven participants, 5 had received a SPEAK score of 45, one received 40, and one received a passing score of 50 during the spring 2009 semester, while enrolled in ESL 504.

Participants ranged in age from 23-28 (average = 25 years), all were Mandarin speakers from mainland China, and all had been in the U.S. for 9 – 33 months (average = 16). Average years of instruction was 10 (range = 6 – 13). One student (Vicky) had taken ESL 504 once previously, though with an instructor who did not use the targeted strategies; William and Jeff had taken a

UIUC course (ESL 506) in which pronunciation instruction was a minor component. Jeff had used listening, transcription, and rehearsal five or fewer times during ESL 506; William used listening and rehearsal during three instructor office visits when taking ESL 506. Participants' demographic data are summarized in Table 1.

Table 1. Demographic Data for the Seven Study Participants

Participant	Sex	Age	Home country	Discipline	Months in US	Prior pronunciation instruction?	SPEAK score	Prior strategy use?	Years of English instruction
Andrew	Male	26	China	Computer science	10	no	45 Spring 08	no	13
Jeff	Male	28	China	Statistics	33	UIUC Fall 2008 (minimal focus)	45 Spring 08	All 3 strategies, 5 or fewer times, Fall 2008	11
Nancy	Female	25	China	Statistics	33	no	45 Spring 08	No	6
Vicky	Female	24	China	Sociology	9	Same class Spring 2008	50 Spring 09	No	10
Wendy	Female	23	China	Engineering	9	no	40 Fall 08	Listened to recording 5 or fewer times for TOEFL prep; rehearsal used for prep for presentation	7
William	Male	25	China	Biophysics	9	UIUC Fall 2008 (minimal focus)	45 Spring 08	Listen to recording, rehearsal during office hours with teacher	unknown
Yvonne	Female	24	China	Biology	9	No	45 Fall 08	No	10
Average		25			16				10

### **Strategy and Pronunciation Instruction**

Students received self-monitoring strategy training and instruction on English suprasegmentals throughout the semester (suprasegmental instruction was based on the course text, *Speechcraft* (Hahn & Dickerson, 1999). The instructor provided feedback in several ways: during class (twice weekly sessions of 80 minutes each), as written or audio (recorded) feedback on strategy use and pronunciation following weekly homework assignments, and during individual 20-minute meetings with each student, held three times during the semester (following each of three mini-lecture presentations) to review progress on target features and strategy use.

### **Materials**

Speech samples were gathered from students' third 5-minute mini-lecture (ML), presented during the final week of the semester. Mini-lecture topics represented content from students' fields of study. Students were instructed not to memorize the text nor were they allowed to read from a written text. The last portion of each speech sample typically included the student's responses to audience questions.

Each mini-lecture was audio-recorded in the classroom, using a cordless microphone and a Sony digital recorder. Each mini-lecture was divided by the researcher into six separate but equivalent audio files. These recordings were used by the students to complete the experimental tasks. The MU was chosen as the unit of analysis because the target pronunciation features were suprasegmentals, for which the MU is the relevant environment for analysis. Students in ESL 504 are instructed to produce message units that are approximately five to nine words in length, following research on short-term memory originated by Miller (1956), suggesting a limit of seven plus or minus two chunks of information. Message unit boundaries typically, but do not always, match boundaries of grammatical units such as noun or verb phrases or clauses. Following is an excerpt from a student's mini-lecture, showing the MUs the student produced and also a "target version", or what they would be expected to produce, based on instruction provided during the semester:

ML: so / given the situation that everybody know, / for example if you / if you are / having a / a critical test. /

Target version: so / given the situation that everybody know, / for example / if you are having a critical test. /

Ten days after the final class, students completed the experimental tasks in a lab, using computers equipped with headphones for listening and microphones for recording. Additional materials included written instructions, checklists to guide task completion, and paper and pencils for writing down and annotating (marking up) transcriptions.

**Procedure**

Due to institutional constraints (i.e., limited class time for completing the tasks and the need for participant identities to remain anonymous to the researcher), experimental tasks were completed on one day, during a 2 to 2-1/2 hour session with regular breaks. Tasks were completed in the following order: L, LT, LTA. For each participant, early, middle, and late speech segments were randomly assigned to the three self-monitoring tasks. This was done to prevent systematic bias based on location of a segment in the speech sample. However, for each experimental task, the speech segments were presented in the order in which the participant originally presented them, so that their production of the suprasegmental features would reflect the original discourse structure.

For the L task, participants listened five times to a segment representing approximately one sixth of their lecture (15 MUs). They were instructed to listen to one suprasegmental feature each time. Next, they were presented aurally with a 1- to 2-MU portion of the larger segment and were told to “Listen and repeat one time”. This listening and repetition (rehearsal) phase occurred three times for each portion until each 15-MU segment was completed. Each rehearsal was audio-recorded. This process was repeated for the next 15-MU segment from a later portion of the mini-lecture.

For the LT task, participants followed a similar process of listening to two different segments of their lecture. During the first listening, participants transcribed the segment. During subsequent listenings, they focused separately on each targeted suprasegmental feature.<sup>1</sup> After completing the transcription, participants read and orally corrected the transcribed segment a total of three times. Each reading was audio-recorded.

For the LTA task, participants completed the listening and transcription steps as described for the LT task (using two new sections of the lecture), then systematically reviewed the transcriptions for non-target pronunciation, marked corrections (annotated) in a different colored pencil, and then read and orally corrected each segment a total of three times. Each reading was audio-recorded. Figure 1 summarizes the experimental procedures.

---

1. Students present and record ML during class (final three class sessions of the semester)			<b>% A C C U R A C Y</b>
2. Ten days after the ML, participants use each self-monitoring combination for different segments of the ML.			
Two 15-MU segments	Two 15-MU segments	Two 15-MU segments	
<b>Listening</b> ↓	<b>Listening + Transcription</b> ↓	<b>Listening + Transcription + Annotate corrections</b> ↓	
Oral rehearsal x 3	Oral rehearsal x 3	Oral rehearsal x 3	
3. Researcher transcribes the ML and 1 <sup>st</sup> and 3 <sup>rd</sup> rehearsals and compares accuracy for each to a target version. These two post-monitoring accuracy scores will be compared to the accuracy score for the original ML.			

Figure 1. Summary of experimental procedures.

### Preparation of data for analysis

Each participant's mini-lecture recording was transcribed by the researcher, and MU boundaries, PPS, and intonation patterns were noted. The researcher prepared a "target" version of the mini-lecture text (Figure 2), showing the suprasegmental features the students would be expected to produce correctly, based on the semester's instruction. The researcher transcribed participants' first and third oral rehearsals, marking the targeted suprasegmental features. The following transcription conventions were used:

MU boundaries are marked with a forward slash (/).

PPS is marked using a solid black dot (●) above the syllable receiving PPS.

When a syllable that should be unstressed is given a heavy stress (but not PPS), an open circle (○) is placed above the syllable. This often occurred on pronouns that should have been unstressed. Intonation patterns were noted as follows: a comma (,) denotes a rise to mid- or



high-range (non-final or question intonation); a period ( . ) denotes a fall to low range, or phrase-final intonation.

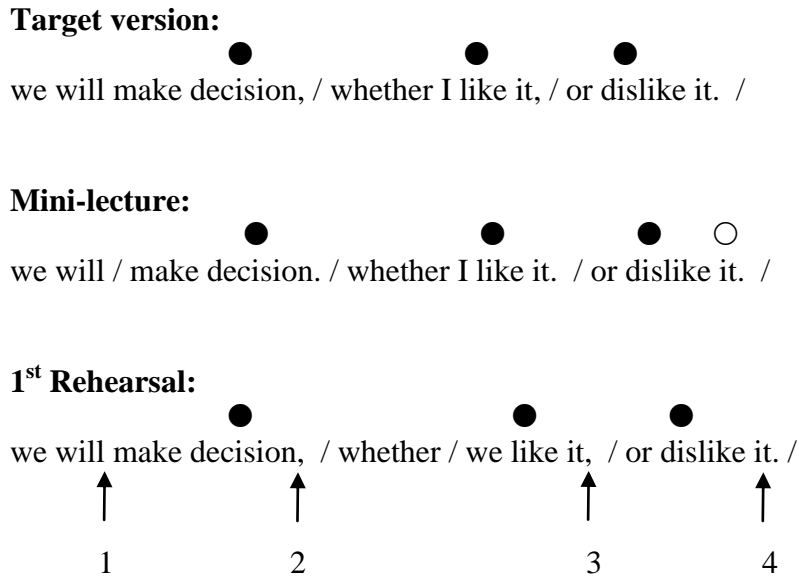


Figure 2. Data sample for Jeff: 1) deleting unnecessary MU break; 2) and 3) correcting intonation from final to non-final; and 4) de-stressing function word (“it”) to highlight PPS.

### Data Analysis

**Selection of equivalent speech samples.** Message units were eliminated from the data analysis if the speech produced was not equivalent across all versions (i.e., mini-lecture and 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> rehearsals). For example, if a participant added or deleted text in one version or revised the wording such that a particular MU was no longer parallel in content and phonological structure, that MU was deleted from the analysis. After unacceptable MUs were omitted, a total of 24 MUs per strategy combination could be used from each participant (from an original 30 MUs).

**Accuracy scores.** During data analysis, MU boundaries, PPS, and intonation were marked as correct or incorrect, based on a comparison with what the student could be expected to do following the course’s instruction. The correct targets were tallied to result in an accuracy score (represented as a percentage of correct targets) for each strategy combination used. The accuracy score for the mini-lecture was considered the baseline level for each participant (what they could do without self-monitoring). Accuracy scores for the first oral rehearsal were used to determine the extent a specific strategy combination resulted in increased accuracy for each pronunciation target. Scores for the third oral rehearsal provided evidence for the effectiveness of repeated rehearsals.

Accuracy scores were calculated for each of the following data categories: (a) by strategy combination for all three suprasegmentals combined: for the total group, and separately for each individual; (b) by strategy combination for each suprasegmental feature: for the group and for each individual.

## RESULTS

**Research question 1: In what ways does strategy use (L, LT, LTA, rehearsal) result in improved pronunciation accuracy? To what extent do strategy combinations have differential effects?**

**Group Results.** Overall group results (Table 2) show that accuracy scores for the three pronunciation targets improved for each type of self-monitoring, from 76% to 82% for L, 76% to 81% for LT, and from 75% to 83% for LTA. The sample size did not permit tests of statistical significance for these differences. Grouping the data masks individual differences in strategy effectiveness, thus individual results are provided next.

*Table 2. Group Accuracy Scores and Percentage Difference Following the Use of Self-Monitoring and Rehearsal*

<b>All Participants</b>	<b>Mini-lecture 3</b>	<b>1<sup>st</sup> Rehearsal</b>	<b>% difference</b>
Across all conditions	76%	82%	+6%
Listening only	76%	82%	+6%
Listening + Transcription	76%	81%	+5%
Listening + Transcription + Annotation	75%	83%	+8%

**Individual results.** Individual results for each strategy combination, regardless of pronunciation target, appear in Figure 3. L was the most effective strategy for three participants (Nancy, 11%; Vicky, 7%; Wendy, 13%) and was also highly effective for a fourth (Yvonne, 11%); LT was most effective for only one (Andrew, 11%); and LTA was highest for William (11%) and Yvonne (14%), and LT and LTA were both most effective for Jeff (10%). Thus L and LTA resulted in the most frequent improvement. These individual findings are consistent with the group results showing greater effects for LTA and L as compared to LT.

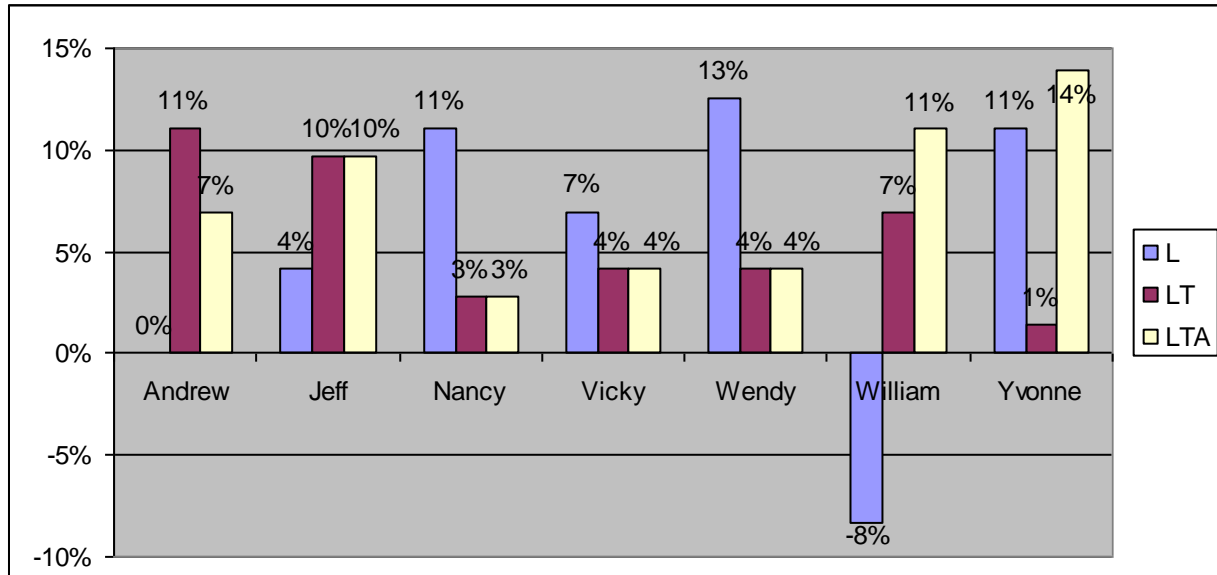


Figure 3. Percentage improvement for each participant, by strategy combination.

**Effects of Rehearsal.** When looking at group average results, regardless of strategy type, the greatest improvement following self-monitoring occurred with the first rehearsal, with a difference of 7% for MUs and PPS, and 6% for Intonation (Table 3, Figure 4). As expected, accuracy gains were not as strong between the first and third rehearsal; there simply is less room for improvement following subsequent rehearsals. However, when looking at each strategy type, we see that accuracy following the third rehearsal was the same as or lower than R1 in several instances, including use of LTA for message units, L and LT for PPS, and LT and LTA for Intonation.

Table 3. Effects of Rehearsal

	Message units			Primary Phrase Stress			Intonation		
	ML	R1	R3	ML	R1	R3	ML	R1	R3
L only	82%	86%	88%	65%	75%	74%	85%	87%	88%
LT	80%	86%	90%	65%	68%	68%	81%	89%	86%
LTA	83%	92%	90%	61%	67%	70%	81%	88%	85%
Average	81%	88%	89%	63%	70%	71%	82%	88%	86%

Note: R1 = first rehearsal; R3 = third rehearsal

Because the accuracy score gains were greatest for the first rehearsal and scores changed minimally or slightly declined at R3, the remaining data analysis focused on the change from the mini-lecture to the 1<sup>st</sup> rehearsal.

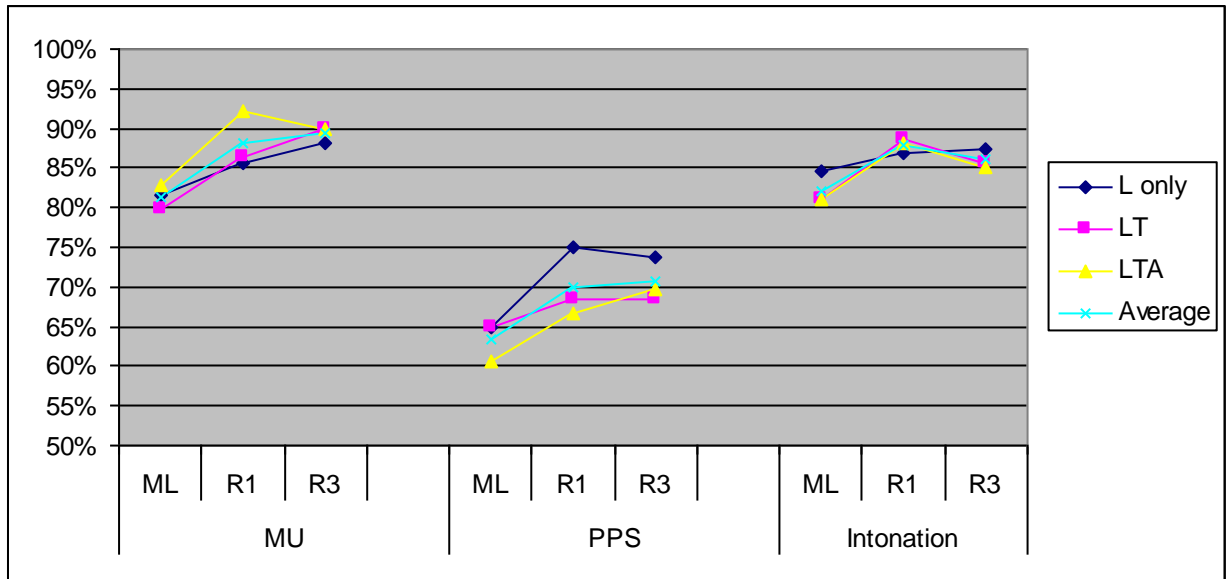


Figure 4. Change in accuracy scores (group values) for the mini-lecture (ML), 1<sup>st</sup> rehearsal (R1), and 3<sup>rd</sup> rehearsal (R3), for each pronunciation target, by strategy combination.

**Research question 2. Effect on MUs, PPS, Intonation Accuracy - How effective are the strategy combinations for each pronunciation target?**

When looking at group results by pronunciation target (Table 4), the difference in accuracy from the mini-lecture to the first rehearsal was greatest for PPS (7%), followed by MUs (6.7%), and then Intonation (6%). These findings pattern with the beginning accuracy score for each target, with PPS the lowest (65%), followed by MUs (82%), and Intonation starting at the highest accuracy level (85%). Thus PPS had the greatest room for improvement.

Table 4. Group Values for Percentage Difference Between Accuracy at the Mini-Lecture and the 1<sup>st</sup> Rehearsal, by Target.

	% difference
PPS	7.0%
MUs	6.7%
Intonation	6.0%

When looking at each pronunciation target, results indicate accuracy increases were greatest for MUs when using LTA (9.4%), followed by LT (6.4%), and L (4.3%). The pattern differed for PPS, with L (10.4%) > LTA (6%) > LT (4.4%). And for Intonation, yet a different pattern emerged: LT (8%) > LTA (7.3%) > L (2.6%) (See Figure 5).

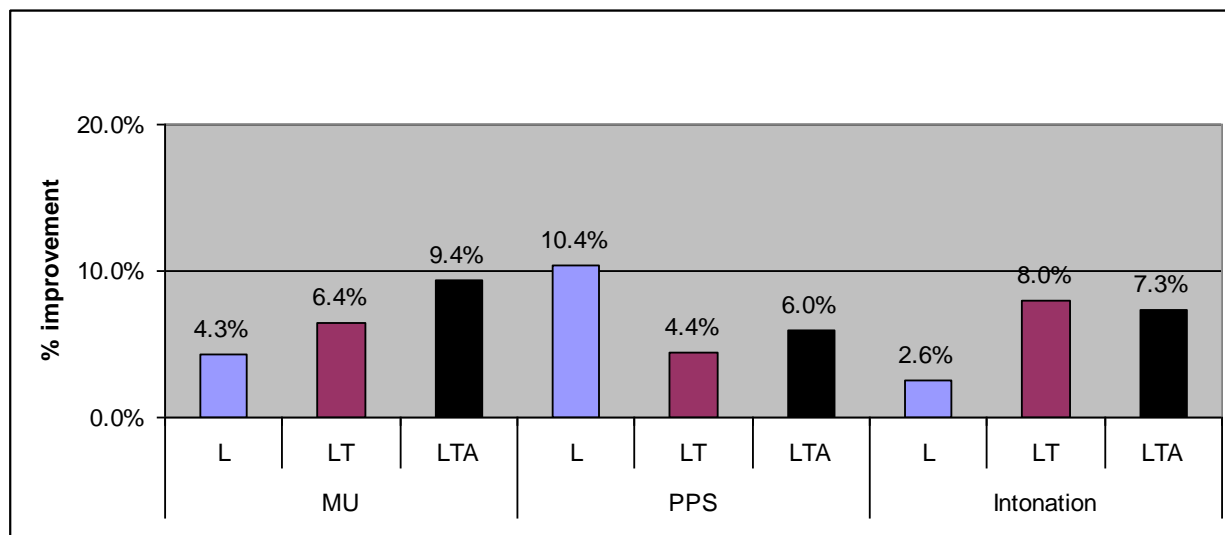


Figure 5. Percentage improvement from mini-lecture to 1<sup>st</sup> rehearsal for the group, by pronunciation target and strategy combination.

**Individual Results for Pronunciation Targets.** Individual performance mirrors group results for each target. Use of one of the transcription combinations (LTA or LT) was beneficial for most participants, with the exception of PPS. LTA was most successful for MUs. L was the most useful strategy for PPS, and LT was most useful for Intonation (Table 5).

Table 5. Breakdown of Effective Strategy Combinations by Target, by Participant. (Totals May Exceed 7.)

Target	Strategy Combination		
	L	LT	LTA
MUs	3	2	5
PPS	5	2	1
Intonation	1	4	3

## DISCUSSION

Overall, the use of self-monitoring strategies results in increased pronunciation accuracy scores and specific strategy combinations have different effects for each of the three pronunciation targets.

**Research question 1: In what ways does strategy use (L, LT, LTA, rehearsal) result in improved pronunciation accuracy? To what extent do strategy combinations have differential effects?**

Group results suggest that the LTA combination is most effective for this group of L2 learners (Table 2). These findings are consistent with those of Mennim (2003, 2007) and suggest that the

use of transcription may help learners notice, focus on, and identify pronunciation features that are more difficult to detect when listening holistically to one's speech (Izumi, 2003; Schmidt, 1993). Pennington and Ellis (2001) also concluded that unless L2 speakers' attention is drawn to the functioning of prosody in the L2, they likely won't attend to it, especially when the role of prosody is different in their L1 (as is the case with Mandarin). Thus training is necessary to help L2 learners, even at an advanced stage, to notice and interpret suprasegmental features. Following are other factors important in the interpretation of the data.

***Prior Use of Strategies.*** Three participants indicated prior use of self-monitoring strategies. Jeff used all three combinations during another UIUC ESL course, most likely following three presentations. Wendy used listening and rehearsal for preparation for a presentation and for the TOEFL test. William indicated that he used listening and rehearsal during office hour visits with his instructor (typically 3 to 4 times per semester). In each case the strategies were used 5 or fewer times, which makes it unlikely that these participants received sufficient practice to influence their performance in this study.

***Possibility of Ceiling Effects.*** Beginning accuracy levels for several participants exceeded 90% for MUs (Jeff, William, and Wendy) and Intonation (Nancy, William), suggesting that these participants had nearly mastered these features and had little room for gains in accuracy. Thus these features were not the best choice for this study for these participants. However, the story for PPS is different. All participants' initial accuracy scores for PPS started at or below 83%, with the lowest level at 42% (Figure 4), so ceiling effects are not likely operating for this pronunciation target. Analysis of other target features in the remaining speech data, including linking, vowel reduction, and word and compound noun stress, may, like PPS, start with lower accuracy levels and thus offer clearer insights into strategy effects. L2 instructors will want to rely on results of diagnostic tests to determine the targets for which learners will derive the greatest benefit from self-monitoring.

***Effects of Task Load.*** An unexpected finding from this study was that multiple rehearsals did not necessarily result in improved accuracy scores. Group results showed that differences in R1 and R3 accuracy were only +1% for L and LT and -2% for LTA. Several factors may help explain this pattern.

- **Fatigue.** Participants completed the tasks during one session of 2.5 hours or less. This was done to accommodate the course schedule, which limited the amount of class time available for completing the experimental tasks. Ideally, the tasks should be completed in shorter sessions over several days, and in actual practice, this is indeed what L2 learners would do. LTA was always the final task, which would lead one to expect the greatest effects of fatigue when using this strategy combination. However, when looking at individual performance across the three tasks, accuracy levels for only 2 of the 5 participants declined over time (from L to LTA) and only for Intonation (Nancy and Wendy). Though I did not track the

amount of time that each strategy combination required, I observed that participants took longest to complete the Listening only task. Except for PPS, this strategy was least effective for most learners, thus some participants may have felt boredom or frustration over the length of the task.

- **Lack of motivation to complete the task accurately.** Most students in ESL 504 failed the SPEAK test and are in the class because it is a requirement to retake the test. Thus one might expect students' intrinsic motivation levels to be low. Motivation levels were not directly targeted in this study. However, on a post-experiment questionnaire, six of seven participants rated the usefulness of Listening and Rehearsing higher than transcription. This might suggest that participants should perform better on subsequent rehearsals regardless of task, though the results do not support this.
- **Task effects.** Across the three rehearsals in the LT and LTA tasks, learners may have started to rush and focus less on producing accurate target features during R2 and R3. The findings from the current study conflict with those of Lynch and McLean (2001), who found that repetition resulted in pronunciation improvements, due to task familiarity. Further investigation is needed to better understand the current findings.
- **Memory effects.** Learners may be over-taxing short-term and working memory as they process multiple chunks of text during rehearsal.
- **Second guessing.** By consciously attending to L2 features that they previously produced automatically, participants may be doubting their initial choices and "correcting" something that was accurate to begin with (Willingham & Goedert-Eschmann, 1999). If second-guessing is a factor, one would expect greater declines on MUs and Intonation, given their high beginning accuracy levels. As noted previously, this did happen for Intonation for Nancy and Wendy, but not for other participants.

Certainly the current study's preliminary findings are not sufficient to contradict long-held beliefs about the importance of rehearsal for improving L2 pronunciation. However, the findings do suggest that L2 learners may need additional training on how to maximize benefits from rehearsal and how to maintain concentration on the task.

**Individual Differences.** Participant performance was variable across the three strategy combinations (Figure 3). For example, Andrew was most successful using LT and showed no change in accuracy when using L. Conversely, Nancy, Vicky, and Wendy were most successful using L. These seemingly inconsistent results at the individual level may be due to differences in learning styles (e.g., visual vs. aural styles), second-guessing previous decisions, or incorrect

---

application of rules. Participant interviews and analysis of transcripts from the LT and LTA tasks may provide insights into possible learner preferences for a particular strategy combination.

**Research question 2. Effect on MUs, PPS, Intonation Accuracy - How effective are the strategy combinations for each pronunciation target?**

Results suggest that learners may find greater success when using LTA for MUs, LT for Intonation, and L for PPS (Figure 5). What are possible reasons for these differences? First, participants had already achieved a high level of accuracy for MUs and Intonation and thus may be better equipped to monitor these targets successfully. Perhaps MU boundaries and intonation contours are easier to “see” in a transcription, and harder to remember when listening only. That is, the visual cues of slashes ( / ) for MU boundaries and arrows for intonation markings ( ↑↓ ) may be more memorable and easier to implement during rehearsal. Perhaps when reading, learners do not need to rely on memory and can use the visual cues of the transcript as reminders. Regarding PPS, perhaps PPS is more salient for learners when presented *aurally* rather than *visually*. Seeing the dot ( ● ) over a syllable receiving PPS may not be sufficiently meaningful, but, during listening and speaking, learners may be able to hear and feel the difference between target-like and non-target-like PPS. As noted earlier, at least for PPS, analysis of a transcript may lead to more second-guessing than listening only. Further analysis of learner transcriptions and comparisons of their predictions and performance in the LTA condition are needed to better answer this question.

**Limitations**

Several limitations of this study are evident and offer potential for future research. First, the sample size is small. These preliminary findings will be expanded as the data for the remaining eight participants are analyzed. Second, additional data are needed to determine the efficacy of self-monitoring for learners at low and intermediate levels. Third, inconsistencies in task types may influence outcomes. Some tasks required reading, others listening only. The original mini-lecture was presented to an audience, however, the experimental tasks were not. Though the focus of this study is on strategies for self-study, an interesting question is to what extent interaction may result in more accurate production when using self-monitoring.<sup>2</sup> Further work is needed to understand how such task differences may affect the study’s results. Fourth, R2 was not analyzed, making it unclear how accuracy scores evolved from R1 to R3. R2 effects will be investigated during the next data analysis stage. Fifth, the results don’t offer insights into long-term effects of self-monitoring. And lastly, the tasks used in this study were not authentic examples of covert rehearsal. Rather than completing tasks in privacy, participants completed the tasks in a computer lab that resembled a testing situation. However, in order to ensure tasks were completed in a consistent manner using equivalent speech samples, the artificiality of a semi-experimental setting was necessary.



## CONCLUSIONS

The results of this study move us a few steps further in our understanding of the merits of strategy use for L2 pronunciation improvement. Prior research has looked at global pronunciation change (e.g., Acton, 1984; Mennim, 2003) and has not tried to isolate the effects of self-monitoring strategy use on message unit boundaries, primary phrase stress, and intonation accuracy. Knowing that specific self-monitoring strategy combinations may be more suitable for specific pronunciation targets can aid language teachers as they target their instruction. Self-monitoring skills may be useful for learners interested in enhancing their study practices for traditional and online classes and for post-instruction study.

Completion of the data analysis for this study will further illuminate the efficacy of self-monitoring strategies for L2 pronunciation and the relationship between strategy types and pronunciation targets. Further research is needed to explore long-term benefits of self-monitoring strategy use and further expand our understanding of how language teachers can facilitate L2 pronunciation improvement for their students.

## NOTES

1. The full study will analyze the participants' use of rhythm facilitators and word and construction stress. These targets are not included in this paper.
2. I would like to thank Colleen Meyers for pointing out the potential value of interaction in making this study's pedagogical tasks more realistic and the possibility that the communicative component may further enhance self-monitoring accuracy.

## ABOUT THE AUTHOR

Sue Ingels, MATESL, is a PhD candidate in the department of Educational Psychology at the University of Illinois at Urbana-Champaign. Her ESL teaching experience includes courses in pronunciation, business communication, academic and business writing, and international teaching assistant training, all at UIUC. She also has coordinated the MATESL teaching practicum. Her research interests include acquisition of English suprasegmentals by ESL learners, L2 phonological acquisition and pronunciation instruction, and teacher education. Email: [ingels@illinois.edu](mailto:ingels@illinois.edu)

## REFERENCES

- Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly*, 18(1), 71-85.
- Benrabah, M. (1997). Word-stress: A source of unintelligibility in English. *International Review of Applied Linguistics*, 35, 157-165.
-

- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.
- Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, 18(3), 53-70
- Cruttenden, A. (1997). *Intonation* (2<sup>nd</sup> ed.). Cambridge, UK: Cambridge University Press.
- Dickerson, W. B. (1987). Explicit rules and developing interlanguage phonology. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 121-140). Berlin: Foris Publications.
- Dickerson, W. B. (1989). *Stress in the speech stream: The rhythm of spoken English* [CD-ROM]. Urbana, IL: Author.
- Dickerson, W. B. (1994). Empowering students with predictive skills. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 19-35). Arlington, VA: TESOL.
- Dickerson, W. B. (2000). Covert rehearsal as a bridge to accurate fluency. Presented at the 34<sup>th</sup> Annual Meeting of the Teachers of English to Speakers of Other Languages, Vancouver, BC, Canada, March 18, 2000.
- Eckstein, G. T. (2007). *A correlation of pronunciation learning strategies with spontaneous English pronunciation of adult ESL learners*. Unpublished master's thesis. Brigham Young University.
- Gu, P. Y. (2007). Foreword. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. vii-viii). Oxford: Oxford University Press.
- Guion, S. G., Harada, T., & Clark, J. J. (2004). Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition*, 7, 207-226.
- Hahn, L. D. (1999). *Native speakers' reactions to non-native speech in English discourse*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- 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. D. (1999). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor, MI: University of Michigan.

- Hsiao, T.-Y., & Oxford, R. L. (2002). Comparing theories of language learning strategies: A confirmatory factor analysis. *The Modern Language Journal*, 86(3), 368-383.
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, 24(2), 168-196.
- Juffs, A., & Rodriguez, G. (2007). Working memory in context: Differential effects on comprehension of relative clauses and anaphor binding. Presented at Second Language Research Forum, Urbana, IL, October 13, 2007.
- Kawagoe, I. (2003). Acquisition of English word stress by Japanese learners. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th generative approaches to second language acquisition conference* (pp. 161-167). Somerville, MA: Cascadilla.
- Lightbown, P. M., & Spada, N. (2006). *How languages are learned* (3<sup>rd</sup> ed.). Oxford: Oxford University.
- Lynch, T. (2007). Learning from the transcripts of an oral communication task. *ELT Journal*, 61(4), 311-320.
- Lynch, T., & Maclean, J. (2001). 'A case of exercising': Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 141-162). Harlow: Longman.
- Mennim, P. (2003). Rehearsed oral L2 output and reactive focus on form. *ELT Journal*, 57(2), 130-138.
- Mennim, P. (2007). Long-term effects of noticing on oral output. *Language Teaching Research*, 11(3), 265-280.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2), 81-97.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25(3), 481-520.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. New York: Newbury House.
- Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, 84, 372-389.
-

- Peterson, S. S. (1997). *Pronunciation learning strategies and learning strategies related to pronunciation ability in American university students studying Spanish*. Unpublished doctoral dissertation, The Ohio State University.
- Peterson, S. S. (2000). *Pronunciation learning strategies: A first look*. Unpublished research report. (ERIC Document Reproduction Service ED 450 599; FL 026 618).
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35(2), 233-255.
- Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*, 20, 213-243.
- Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly*, 9(1), 41-51.
- Rubin, J. (1981). Study of cognitive processes in second language learning. *Applied Linguistics*, 11(2), 118-131.
- Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, 13, 206-226.
- Stern, H. H. (1975). What can we learn from the good language learner? *The Canadian Modern Language Review*, 31(4), 304-318.
- Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal*, 90, 519-535.
- Tyler, A., Jeffries, A., & Davies, C. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teachers' spoken discourse. *World Englishes*, 7(2), 101-110.
- Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly*, 24(2), 177-198.
- Vitnova, G., & Miller, A. (2002). Reflective practice in pronunciation learning. *The Internet TESL Journal*, VIII(1), April 24, 2008.
- Wennerstrom, A. (1998). Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition*, 20(1), 1-25.
- Willingham, D. B., & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological Science*, (10)6, 531-534.

## **A Longitudinal Investigation of Vowel Acquisition**

**Fran Gulinello**

Nassau Community College, Garden City

A longitudinal study investigated change in the vowel systems of five adult native Spanish speakers learning English. It focused on 11 vowels of English as uttered in CVC words and in various sentential contexts. Vowel productions from each speaker were measured for the acoustic parameters of F1, F2 and duration. These acoustic parameters were then analyzed via classification matrices of discriminant analysis and compared over time. Findings indicate that the vowels of nonnative speakers change in ways that reflect dialectal and diachronic changes. Specifically, we see instances of split, merger and shift as described by Labov (1994). It is also the case, however, that changes occur that are unique to second language (L2) acquisition. This study provides evidence that the intermediate phonological systems arising during L2 acquisition should be viewed not only in terms of the target but as unique systems of contrasts. It also provides evidence that changes are not necessarily unilateral; movement in one area of the system can affect other areas of the system. These findings are relevant to the way in which we view, teach and assess the pronunciation of an L2 vowel system.

### **INTRODUCTION**

In the study of second language (L2) acquisition and cross-linguistic production and perception, many models have been proposed to account for learners' pronunciation of L2 sounds. In an attempt to determine which model best accounts for the complex facts, it is often the case that we focus on particular aspects of pronunciation, on particular subsets of the phoneme inventory, or on particular learning paradigms that will best distinguish between models or hypotheses. The methodology for studying L2 phonological acquisition is often a three-step process: 1) form a hypothesis, 2) find a subset of the phonological system that has characteristics to test the hypothesis, and 3) determine whether the actual productions by L2 learners on that subset support the hypothesis. The results of such studies have undoubtedly brought us closer to understanding L2 phonological acquisition in terms of both production and perception. Yet, we are left with a crucial question. Do all the sounds or sound contrasts that fall into the same production or perceptual pattern behave the same way within an individual speaker? In her study of cross-linguistic perception, Escudero (2000) suggests several directions for future research. She specifically mentions the need to study category formation and perception of other sounds in the system. She also suggests the need for longitudinal studies of one year or more to determine the stability of the patterns and the sequence of development. These suggestions are clearly relevant to L2 production as well.

This study examines the development of phonological systems or subsystems over time. It is a descriptive study of change in the non-diphthongized vowel systems of L2 learners. Its purpose

---

is to observe and describe change from two perspectives. The first perspective is that all change is relevant regardless of its nature. This perspective is based on the concept of Interlanguage (IL). Since interlanguage is the unique system of a learner unlike the native or target language (Selinker, 1972), it is possible that the IL can change but still not approximate the target. Thus, in this study, the learners' vowels were first described independently of the target via a comparison of a system at one point in time to *itself* at another point in time. The second perspective of change is based on the idea that successful phonological acquisition entails movement of the system towards the target norms. Thus, change was also observed via a direct comparison of a learner's system to the target system. Similarities and differences between each learner's system and the target system were then compared across time.

## **METHODS**

Participants for this study consisted of five nonnative speakers and two native speakers of English. The five non-native speakers all spoke Spanish as their native language. Three of them were from Colombia, one was from Guatemala, and one was from Peru. Their ages ranged from 20 to 42 years old, and length of residency at the onset of the study ranged from six months to twelve years. Four of the non-native speakers were male and one was female. All of the non-native speakers were enrolled in at least one English as a Second Language (ESL) class at Nassau Community College at the time their first recordings were taken, but only two were registered in a pronunciation class.

The native speakers consisted of one male speaker and one female speaker. At the onset of the study their ages were 34 and 38 years old respectively. They were both born in New York and had lived there all of their lives. They were monolingual English speakers. Both studied Spanish in High School, but neither could converse in any language other than English. The native speakers were used to provide a baseline to evaluate change over time of the non-native speakers.

### **Procedures and Materials**

The portion of the study reported here included a sentence reading task. This task was designed to elicit eleven target vowels: [i, ɪ, e, ɛ, æ, ʌ, u, ʊ, o, ɔ, ɑ]. Target vowels appeared in three monosyllabic English words of the structure CVC, where each C was a stop (Table 1). Each monosyllabic word was then repeated five times with each occurrence of a word being in a unique sentential context.

*Table 1. Target Words*

[i] - <i>keep, peek, bead</i>	[ɪ] - <i>pick, pit, kid</i>
[e] - <i>take, gate, paid</i>	[ɛ] - <i>get, pet, bed</i>
[æ] - <i>cat, pat, bad</i>	[ʌ] - <i>cut, cup, but</i>
[u] - <i>boot, coop, tube</i>	[ʊ] - <i>put, took, could</i>
[o] - <i>coat, boat, code</i>	[ɔ] - <i>talk, caught, taught</i>
[ɑ] - <i>pot, cop, top</i>	

The entire data set resulted in 15 utterances of each vowel<sup>1</sup> (three words multiplied by five repetitions of each word) and a total of 165 different sentences (15 tokens of each vowel multiplied by 11 vowels).

#### *Acoustic Measurement and Statistical Analysis*

Vowels were measured for the acoustic parameters F1, F2, and duration. Duration measures for these CVC words were taken from the release of the first stop gap to the closure of the final stop gap. Formants were measured at the 25%, 50% and 75% points. This means that the duration of the vowel was calculated and then frequency measurements were taken one quarter of the way into the vowel (i.e. the 25% point), one half of the way into the vowel (i.e. the 50% point) and three quarters of the way into the vowel (i.e. the 75% point). This paper focuses on the F1 and F2 measurements from the 50% point because these were clear across all speakers and were highly consistent when random samples of the data were re-measured.

Acoustic measurements were analyzed via discriminant analysis; this is a multiple regression technique that examines a set of variables or predictors that serve to distinguish a set of categories. Independent parameter values are weighted to maximally distinguish separate categories. For this investigation, individual speakers were treated as separate case studies. The 11 *intended* vowels (American English vowels designated for the lexical items) served as the categories; parameter values (F1, F2 and duration) of the vowels *actually uttered* served as input to the model. The discriminant analysis essentially took the acoustic parameters of all vowels entered for an intended group and found a centroid for that group. It then determined how near or far the acoustic parameters for each uttered vowel were from that centroid. If the acoustic parameters of a spoken vowel were close to the centroid of the intended category, the discriminant analysis characterized them as *correct* matches. If the acoustic parameters of a spoken vowel were far from the centroid of the intended category (or closer to the centroid of an unintended category), they were characterized as being mismatched. In this respect, *intended*

---

simply refers to the vowel phoneme that linguists consider to be in the particular word being pronounced. For example, in the word *keep* linguists consider the vowel phoneme to be /i/. Therefore /i/ is the intended category. It does not imply any knowledge on the part of the speaker as to what should be produced in a given English word. The output of the discriminant analysis is a classification matrix as illustrated in Figure 1 and Table 2. Figure 1 shows the F1, F2 plot of a native speaker producing instances of the intended vowel /i/ in the words *keep*, *peek* and *bead*. Table 2 shows how the discriminant analysis categorized each vowel uttered. Notice that 13 of the 15 vowels uttered in the words *keep*, *peak*, *bead* matched the centroid of the vowel /i/. Two instances were closer to the centroid of the vowel /i/. The confusion matrix represented in Table 2 indicates this as 86.6% correct match for the vowel /i/ to its intended category.

Discriminant classifications were used in two types of analyses corresponding to the two perspectives of change discussed earlier. The first analysis considered the vowels of a nonnative speaker as a separate system and did not compare them to any external criteria. This allowed observation of change within a learner's system independent of native speaker norms. The second type of analysis used the centroids of a native speaker's vowels (male to male and female to female) as the criteria for evaluating the vowels uttered by a nonnative speaker. This made it possible to evaluate a learner's system in direct comparison to the target language.

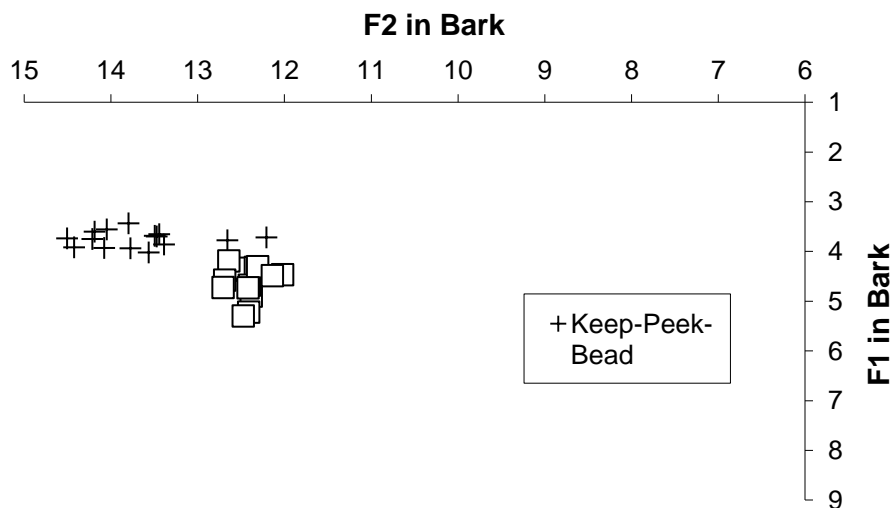


Figure 1. F1/F2 plot of native speaker producing words with /i/ and /ɪ/. The F1/F2 plot of a native speaker of English producing 15 words that contain the vowel /i/ and 15 words that contain the vowel /ɪ/. Note that F1 and F2 measurements have been converted to Bark from Hz. The Bark scale ranges from 1-24 and is a measure of frequency based on the critical bandwidths of hearing.



Table 2. Sample Classification Matrix for /i/ and /ɪ/

Intended Vowel	% correct	Classification				
		i	ɪ	e	ɛ	æ
i	86.66666	13	2	0	0	0
ɪ	73.3	0	11	2	2	0

NOTE: Classification matrix of a native speaker of English producing 15 items that contain the vowel /i/ and 15 items that contain the vowel /ɪ/. The data are the results of a discriminant analysis with F1 and F2 as the only input parameters.

## RESULTS

To begin, we first look at the results of the native speakers. The native speakers in this study were essential for establishing baseline measurements for presenting a realistic picture of the target vowel system on these particular lexical items in these sentences, as evaluated by this particular statistical analysis<sup>2</sup>. The findings regarding the native speakers could be a separate paper, but a brief summary is given below.

1. Native speakers do not have perfect category matches of vowels.
2. Category matches within a single lexical item were better than matches across lexical items.
3. Native speakers exhibit allophonic variation that affects category matches.
4. Native speakers are not completely stable over time.

Evidence for these conclusions can be seen in Table 3 which shows the percent of correct category matches for the first native speaker<sup>3</sup> at Time 1.

Table 3. Overview of Category Matches at Time 1 for a Native Speaker

Percent Matched on F1, F2 and Duration		
Target Vowel	Across Three Lexical Items	In a Single Lexical Item
i	93.3 (14/15)	100 (5/5)
ɪ	33.3 (5/15)	100 (5/5)
e	92.8 (13/14)	100 (5/5)
ɛ	85.7 (12/14)	100 (5/5)
æ	53.3 (8/15)	100 (5/5)
ʌ	53.3 (8/15)	100 (5/5)
u	73.3 (11/15)	100 (5/5)
ʊ	66.6 (10/15)	100 (5/5)
o	60 (9/15)	80 (4/5)
ɔ	66.6 (10/15)	80 (4/5)
ɑ	100 (13/13)	100 (5/5)
<b>Total %</b>	70	96.3

NOTE: Percentage of correct category matches of one native speaker at one point in time with all three parameters entered into the discriminant analyses. This table also compares the percent of category matches of vowels uttered across lexical items as compared to vowels uttered in a single lexical item.

The first column in Table 3 shows the percent correct for each vowel uttered 15 times across three different lexical items (e.g. /i/ in *keep*, *bead* and *peek*; /ɪ/ in *pick*, *bit* and *kid*, and so forth). One can immediately see that, native speakers do not have perfect category matches for any of the vowels. The second column shows the results of the discriminant analyses of each vowel uttered only five times and in only one lexical item (e.g. /i/ in *keep*; /ɪ/ in *pick*, and so forth). The category matches for the vowels are far better when uttered in only one lexical item. Expectations are that native speakers have clearly distinguished, if not perfectly distinguished, vowel groups. In fact, in much of the early research on vowels this is the case. Those studies, however, measured vowels in only one context (as in the classic hVd study by Peterson and Barney, 1952). This study and others like it have shown that the more varied the context, the less discrete the groupings become. One can also see that some vowels showed better category

matches than others. The vowels /i/ and /ɑ/, for example, had consistently high category matches across the native speakers. The vowels /o/ and /ɔ/ had consistently lower category matches across the native speakers<sup>4</sup>. Additionally, the vowels /æ/ and /u/ showed clear allophonic variation, which in turn affected the classification percentages. When the lexical item *bad* was included in the analysis, the percentage of correct matches for the vowel /æ/ was only 53%. This number raised to 100% when *bad* was excluded (i.e. when it occurred in the items *pat* and *cat* only). This is likely due to the New York City pronunciation of raised /æ/ in certain contexts. Likewise, the vowel /u/ in the item *tube* caused a lower percent of correct category matches in this group because of the fronting of /u/ with preceding alveolars (see also Hillenbrand, Clark and Neary, 2001 for similar results). One final note about the native speakers is that they were not perfectly stable over time. Categories did not change by more than four matches across the period of one year in either of the native speakers. More importantly the groupings remained stable in that there was no evidence of vowel groups being added being added or eliminated.

Before turning to observations of change in the nonnative speakers, it should be noted that most of the nonnative speakers initially showed a five-vowel pattern when pronouncing the American English vowels (at least with respect to F1 and F2 plots). One speaker, however, showed few clear vowel groupings. She was the least advanced learner and was most likely struggling with the sentence elicitation task. Although it is not possible to report all of the data here, Table 4 provides an example of the initial category matches for one nonnative speaker. Note that the percent of correct category matches are shown both independently of and in direct comparison to the native speaker.

---

Table 4. Overview of Category Matches at Time 1 for a Nonnative Speaker

Target Vowel	Independently of the Native Speaker		Matched to the Centroids of the Native Speaker	
	Across Three Lexical Items	In a Single Lexical Item	Across Three Lexical Items	In a Single Lexical Item
i	0 (0/10)	60 (3/5)	60 (6/10)	80 (4/5)
ɪ	100 (15/15)	60 (3/5)	20 (3/15)	20 (1/5)
e	73.3 (11/15)	100 (5/5)	80 (12/15)	100 (5/5)
ɛ	64.2 (9/14)	100 (4/4)	35.7 (5/14)	100 (4/4)
æ	73.3 (11/15)	80 (4/5)	66.6 (10/15)	40 (2/5)
ʌ	46.6 (7/15)	60 (3/5)	66.6 (10/15)	40 (2/5)
u	77.7 (9/14)	60 (3/5)	50 (7/14)	60 (3/5)
ʊ	64.2 (9/14)	60 (3/5)	42.8 (6/14)	80 (4/5)
o	86.6 (13/15)	80 (4/5)	13.3 (2/15)	40 (2/5)
ɔ	33.3 (5/15)	60 (3/5)	26.6 (4/15)	0 (0/5)
ɑ	21.4 (3/14)	100 (5/5)	21.4 (3/14)	0 (0/5)
<b>Total %</b>	58	74	44	50

NOTE: Percentage of correct category matches of one nonnative speaker at one point in time with all three parameters entered into the discriminant analyses. This table compares the percent of category matches of vowels uttered across lexical items as compared to vowels uttered in a single lexical item. This table also evaluates the nonnative speaker's vowels independently of and in direct comparison to the native speaker's vowels.

The question of course is whether or not these vowel groups changed over time. Before describing change, two points must be made. First, this study looked at whether the relationship between phonemically relevant pairs changed over time<sup>5</sup>. We could ask, hypothetically, if the percent of category matches for /i/ changed over time, but the answer would be meaningless unless we also looked at how it changed in relation to the other vowels, especially /ɪ/.

The second point is that this study did not begin with a specific hypothesis, but rather was intended to be descriptive. As changes emerged it became immediately apparent that many of them fell into patterns of linguistic change proposed earlier by Labov (1994). Labov referred to the major patterns of linguistic change as split, merger and shift. The next section describes

specific instances of change observed in the nonnative speakers with respect to these major patterns.

### Split

The first type of change observed in the nonnative speakers was *split*. According to Labov (1994), split is the process whereby a preexisting phoneme divides into distinct phonemes. This can occur when two allophones become distinctive upon the loss of a conditioning environment or it can occur when existing word classes divide in what Labov refers to as a *lexical split*. An example of change considered to be split can be seen in a nonnative speaker pronouncing target words containing the vowels /i/ and /ɪ/. This particular split involves the duration parameter and shows change when a nonnative speaker's vowels are analyzed independently of a native speaker's vowels. Table 5 shows that at Time 1, 24 of the 25 words uttered were closest to a single centroid and grouped together as a single vowel. By Time 3, seven of the vowels have become distinct (Figure 2).

The nonnative speaker initially had only one vowel with no distinction between the vowels /i/ and /ɪ/. After one year, the speaker used duration to make the vowels different. The speaker still has not associated the correct word with the correct target vowel.

Table 5. Split by Duration of /i/ and /ɪ/

Target Vowel	Time 1		Time 2		Time 3	
	Match /# of Tokens F1/F2	Match /# of Tokens F1/F2 Duration	Match /# of Tokens F1/F2	Match /# of Tokens F1/F2 Duration	Match /# of Tokens F1/F2	Match /# of Tokens F1/F2 Duration
i N=10	----- ɪ (10)	----- ɪ (10)	----- ɪ (10)	----- ɪ (10)	i (2) ɪ (8)	i (4) ɪ (6)
% correct	0	0	0	0	20	40
ɪ N=15	ɪ (14) ----- e (1)	ɪ (15) ----- -----	ɪ (15) ----- -----	ɪ (14) i (1) -----	ɪ (15) ----- -----	ɪ (12) i (3) -----
% correct	93	100	100	93	100	80

NOTE: Discriminant classifications of a nonnative speaker producing /i/ and /ɪ/. Note that the word *match* refers to how individual vowels uttered matched the centroid of the nonnative speaker's own productions, indicating that this evaluation is independent of the native speaker norms. Columns are divided by time and then subdivided by which input parameters were used.

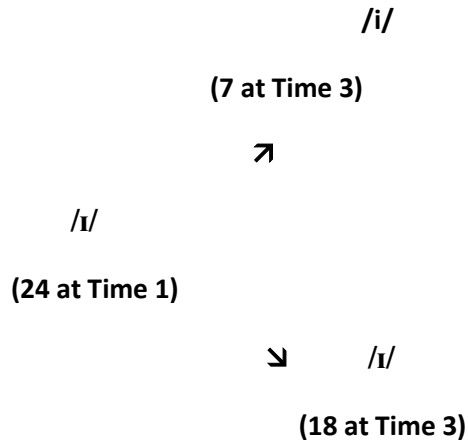


Figure 2. Illustration of split by duration

This case of split is interesting for two reasons. First, it supports the idea of interlanguage, a system unto itself unlike the native language or the target language. These duration differences indicate that a distinction was being learned but it was not the same distinction that the native speakers had. Second, it shows a relationship between perception and production. We know from previous studies that nonnative listeners use duration to distinguish some sounds even when native listeners do not. Bohn (1995) found that while English listeners relied almost entirely on spectral cues to identify the stimuli on the English *beat-bit* continuum, German and Spanish speakers relied heavily on duration cues and Mandarin speakers relied almost exclusively on duration cues (p. 299). This nonnative Spanish speaker relied on duration to split a single vowel grouping of /i/ and /ɪ/ into two vowel groups (whereas the native speakers in the study made a spectral distinction between these vowels). Although we do not have perception data on these particular Spanish speakers, it is interesting to see the similarities between their production and previous perception studies.

Although a few other cases of split were found, it did not occur frequently. Labov (1994) notes that split is a relatively rare linguistic change. Brière (1966) found that *divergence* (when a single sound in the native language must split into two contrastive sounds in the target language) is one of the more difficult patterns to acquire. In this study, split was often accompanied by a merger between another pair thereby preserving the number of original phonemes in the learner's system. Furthermore, the cases of split observed did not always match the native speakers' norms.

## Merger

The second type of change observed in the nonnative speakers was *merger*. According to Labov (1994) merger is the process whereby a phoneme moves in the F1/ F2 space but surrounding vowels do not move. The vowel that is moving essentially encroaches into the space of another vowel and the two become one. Labov contrasts splits and mergers in the following way: Splits involve movement into an *unoccupied* space and *create* distinction; mergers involve movement into an *occupied* space and *eliminate* distinction. Table 6 shows a merger when a nonnative speaker's pronunciation of the target vowel /ʌ/ was compared to the centroid of a native speaker's productions. Notice that at Time 1, the nonnative speaker has some utterances that actually matched the native speaker's centroid for /ʌ/. By Time 3 they have virtually disappeared and almost all match the centroid of the native speaker's /ɑ/.

Table 6. Formant Merger of /ʌ/ and /ɑ/

Target Vowel	Time 1		Time 2		Time 3	
	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2	NS Match /# of Tokens F1/F2 Duration
ʌ	ʌ (4)	ʌ (5)	----	----	----	ʌ (1)
N=9	ɑ (5)	ɑ (4)	ɑ (8)	ɑ (8)	ɑ (9)	ɑ (7)
	----	----	ε (1)	ε (1)	----	ε (1)
% correct	44	55	0	0	0	11

NOTE: Discriminant classifications of a nonnative speaker producing target words with the vowel /ʌ/. NS match refers to how many of the vowels uttered by the nonnative speaker are near the native speaker's centroid for /ʌ/. Columns are divided by time and then subdivided by which input parameters were used.

## Shift

The third type of change observed in the nonnative speakers was shift. Shift refers to a simple movement of a vowel in the F1/F2 space which neither creates nor eliminates phonemic distinction. Table 7 shows an example of shift when the nonnative speaker's production of /e/ and /ɛ/ were compared to a native speaker's centroids.

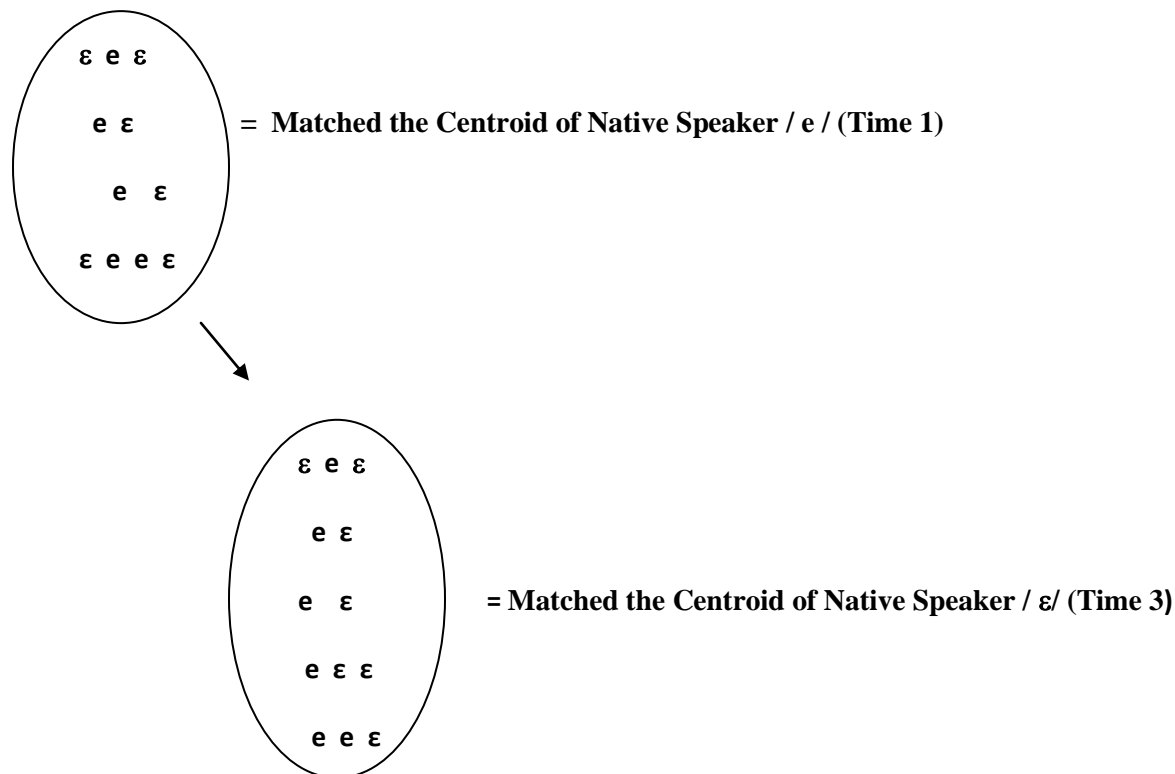
Table 7. Formant Shift of a Mid Front Vowel Grouping

Target Vowel	Time 1		Time 2		Time 3	
	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2 Duration	NS Match /# of Tokens F1/F2 Duration
e	e (9)	e (12)	e (4)	e (11)	e (5)	e (7)
	ε (6)	ε (2)	ε (11)	ε (3)	ε (10)	ε (8)
	-----	ɪ (1)	-----	ɪ (1)	-----	-----
N=15						
% correct	60	80	26.6	73.3	33.3	46.6
ε	ε (5)	ε (5)	ε (12)	ε (11)	ε (13)	ε (10)
	e (7)	e (6)	e (1)	e (2)	e (1)	e (4)
	ɪ (2)	ɪ (3)	ɪ (1)	ɪ (1)	-----	-----
N=14						
% correct	36.7	35.7	85.7	78.5	92.8	71.4

NOTE: Discriminant classifications of a nonnative speaker producing target words with the vowels /e/ and /ε/. *NS match* refers to how many of the vowels uttered by the nonnative speaker are near the native speaker's centroids for /e/ and /ε/. Columns are divided by time and then subdivided by which input parameters were used.

At Time 1, these vowels were poorly distinguished, but the majority of the tokens matched the native speaker's centroid for /e/. By Time 3, the vowels were still poorly distinguished, but the majority matched the native speaker's centroid for /ε/. This change is considered to be shift because a single vowel grouping did not split or merge; rather moved collectively to a different position in the F1, F2 space (as illustrated in Figure 3).





**Figure 3.** Shift of /e/ to /ɛ/. Shift of a mid front vowel grouping. Note that at Time 1, the majority of the utterances matched the native speaker's centroids for /e/. By Time 3, the majority of utterances matched the native speaker's centroid for /ɛ/.

## DISCUSSION

This study has looked at L2 acquisition in a subsystem of the entire phonology as opposed to looking at individual sounds or learning paradigms. It also attempted to describe change over time independently of and in direct comparison to the target language. Findings indicate that change exhibited by L2 learners are in many cases similar to those observed in dialectal and historical change (namely splits, mergers and shifts)<sup>6</sup>.

Additional findings of the study have implications for research and teaching. First, since contrastive pairs that seemingly share characteristics do not exhibit the same types of change, L2 acquisition should not be studied in terms of pairs extracted from the entire system. Likewise, since changes in the system are not always unilateral teachers should use whole system exercises and not minimal pair type exercises alone. This is especially true since there was a general tendency among the speakers in this study to maintain the original number of vowels in the native language system when pronouncing the target language system. Evidence of this was seen in cases where a split in one area was accompanied by a merger in another. Additionally, since

change occurred in the interlanguage system that did not necessarily match native speaker norms, L2 acquisition should not be studied solely with respect of movement towards target but in terms of movement or change in general. Researchers and language teachers should evaluate all learning not just that which achieves the target. Finally, since native speakers overlap, change and do not show perfect category matches for vowels across lexical items, researchers and teachers should acknowledge the actual input as opposed to an idealized input.

## ABOUT THE AUTHOR

Fran Gulinello is an Associate Professor in the Communications Department at Nassau Community College (Garden City, NY). She has a PhD in linguistics from the CUNY Graduate School and University Center in New York. For questions or comments regarding this study, e-mail her at [fran.gulinello@ncc.edu](mailto:fran.gulinello@ncc.edu).

## NOTES

---

<sup>1</sup> Various items were excluded from each speaker's data if they were judged to have been misread, could not be measured, or otherwise not analyzable. All data tables therefore show both percentages and raw scores.

<sup>2</sup> Results of vowel studies vary with respect to the context in which the vowels are couched (Bradlow, 1995; Hillenbrand, Clark and Neary, 2001), the point or points at which the vowels are measured, and the particular dialect of the speakers (Hillenbrand, Getty, Clark & Wheeler, 1995).

<sup>3</sup> For all the data obtained for the native and nonnative speakers please refer to Gulinello (2009).

<sup>4</sup> Perhaps this is due to the New York pronunciation of the vowel /ɔ/ or perhaps other parameters are needed to distinguish these vowels.

<sup>5</sup> A limitation of the study is the incompatibility of methodologies and theoretical frameworks. This study exemplifies such an incompatibility. The methodology collected acoustic data (F1, F2 and duration measurements) and analyzed them via a multiple linear regression technique called discriminant analysis. This method of analysis requires each phoneme to be viewed as an atomic unit devoid of internal structure, a requirement that is in direct conflict with virtually all modern phonological theories. Phonological theories view features, not phonemes, as the atomic units. In particular, this study is an attempt to view the evolution of learner's vowel inventories as changes in a system of contrasts, a perspective which directly entails a featural analysis. The reader should understand that this gap is an artifact of the choice of discriminant analysis as a statistical method, not a theoretical claim on the part of the author. I am grateful to Charles Cairns for pointing this out.

<sup>6</sup> The acquisition of the orthography presents other unique patterns of change which, for space considerations, cannot be addressed here.

## REFERENCES

Bohn, O. S. (1995). Cross language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD: York Press.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, 97(3), 1916-1924.

- Briere, E. J. (1966). An investigation of phonological interference. *Language*, 42(4), 768-796.
- Escudero, P. (2000). *Developmental patterns in the adult acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers*. Unpublished M.Sc thesis, University of Edinburgh, Scotland, UK.
- Gulinello, F. (2009). The acquisition of an L2 vowel system: A longitudinal investigation of change. Unpublished doctoral thesis. City University of New York Graduate School and University Center.
- Hillenbrand, J., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, 109(2), 748-763.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099-3111.
- Labov, W. (1994). *Principles of linguistic change: Internal factors*. Cambridge: Blackwell.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175-184.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10 (3), 209-231.
-

Kang, O. (2010). ESL learners' attitudes toward pronunciation instruction and varieties of English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1<sup>st</sup> Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 105-118), Ames, IA: Iowa State University.

---

## **ESL Learners' Attitudes toward Pronunciation Instruction and Varieties of English**

**Okim Kang**

Northern Arizona University

There is an increasing need for a comprehensive understanding of accent on the part of both instructors and learners. However, researchers in Applied Linguistics have paid little attention to learners' perceptions of pronunciation instruction in L2 contexts. The current study identified adult ESL learners' perspectives of pronunciation studies in the inner circle countries. It reported on students' expectations of their pronunciation lessons and their attitudes toward instructors' accent varieties in the environment of speaking English in New Zealand (NZ) and North America (US). Two hundred thirty eight ESL students participated in interviews and questionnaire surveys. The results of the investigation showed that students in NZ, compared to those in the US, were more dissatisfied with their current curriculum of learning pronunciation due to misunderstanding of various models and accents of pronunciation made available to them. These results suggest that students' perceived needs should be better synchronized in ESL contexts. The implications of this study regarding the relations among pronunciation training, learner perceptions, and accent variety extend beyond the immediate setting in NZ and US higher education, and in fact pertain to the teaching of World Englishes in every nation.

### **INTRODUCTION AND BACKGROUND**

Recently, the role of English as an international language (EIL) has introduced different perspectives to the field of TESOL. The terminology "English as..." refers to proposals that have evolved to describe the increasing amount of communication among speakers that have English as a second language (L2) (Erling, 2005). Modiano (2001), for example, calls EIL an "alternative to standard English", providing a space where speakers can be "culturally, politically, and socially neutral" (p.170). Undoubtedly, Kachru's (1992) model of the spread of English is one of the most influential proposals for the use of English in the world. Kachru divides World Englishes into three concentric circles, the Inner Circle, the Outer Circle, and the Expanding Circle. These three circles represent the types of spread, the patterns of acquisition, and the functional allocation of English in diverse cultural contexts (Jenkins, 2003). The Inner Circle comprises countries where English is historically the first language to be spoken such as the United Kingdom (UK), America, Canada, Australia, and New Zealand. The Outer Circle comprises ESL countries where English has a long history of "institutionalized functions standing as a language of wide and important roles" such as India, South Africa, and Nigeria (Kachru & Nelson, 2001, p. 13). Finally, the English spoken in the Expanding Circle is called English as a foreign language. In this context, English has various roles and is widely studied but

---

for more specific purposes than in the outer circle, including reading knowledge for scientific and technical purpose. Examples of such countries include China, Korean, Iran, Nepal, and Japan.

In light of the internalization of English, the goal of global intelligibility has been emphasized over mastery of a particular native accent (Crystal, 2003; Jenkins, 2000, 2006). In terms of pronunciation teaching, an international version of English, called the ‘lingua franca core,’ has been suggested as a more realistic model. Jenkins (2000, 2002) has argued that learners of English as an international language should not adapt to native speaker (NS) norms but should adjust their speech to suit an audience of primarily nonnative speakers (NNSs). There is also quite wide acceptance that ownership of English no longer belongs just with the inner-circle countries (Kirkpatrick, 2008), given that there are more second language (L2) speakers of English than native speakers (Crystal, 2003; Modiano, 2001; Yano, 2001). In fact, adult L2 learners rarely achieve native-like speech patterns (Moyer, 2004; Scovel, 2000), and native-like pronunciation among those who acquire an L2 after early childhood is difficult to achieve in typical ESL classrooms. Accordingly, scholars such as Derwing and Munro (2005) or Goodwin (2001) argue that teachers should help ESL learners to set realistic goals for pronunciation instruction because if we target native-like accents we may even set our students up for failure.

Among ESL learners, however, there still seems to be a tendency to set inner-circle standards for their own speech. In a study of 100 adult ESL learners in Canada, Derwing (2003) found that the overwhelming majority considered speaking with perfectly native pronunciation to be a desirable goal. Timmis (2002) surveyed around 400 learners among 45 countries and reported that learners preferred to strive for inner-circle norms in their pronunciation. Scales, Wennerstrom, Richard, and Wu’s (2006) study revealed that more than half (62%) of the learners wanted to sound like a native English speaker, even though only 29% were able to correctly identify the American accent.

At the same time, studies have suggested that many learners even prefer to model certain inner circle standards such as Received Pronunciation (RP) or General American (GA). Bayard, Gallois, Ray, Weatherall, & Sullivan (2002) found that students from Europe or Southeast Asia particularly preferred GA. In addition, Bayard’s several other attitudinal studies (1990, 1995, 2001) showed that students rated New Zealand English least favorably, compared to other inner-circle models such as Australian English, GA, and RP. Dalton-Puffer, Kaltenboeck, and Smit’s (1997) study showed that EFL learners performed better and rated RP most highly when they listened to speech samples of both NS and NNS of RP and GA.

Given these trends, questions arise regarding learners’ perspective toward English varieties in inner-circle countries themselves. Which varieties of English should serve as instructional models for pronunciation? Are learners’ reactions still positive if they are exposed to a different variety of inner-circle models? The effects of accent on people’s attitude have been studied for several decades (e.g., Bradac, 1990; Derwing, 2003; Kang & Rubin, 2009; Lippi-Green, 1997; Ryan & Carranza, 1975). These studies have often focused on native speakers’ (NS) biases with

---

nonnative speakers' (NNS) accented speech. Their findings reported that listeners tended to assess NNS accent more negatively than speech that was perceived to be standard. However, researchers in Applied Linguistics have paid little attention to learners' perceptions of pronunciation instruction in ESL contexts in general. A comprehensive understanding of accent varieties on the part of learners is warranted.

This current study reported on adult ESL learners' expectations of their pronunciation lessons and their attitudes toward instructors' accent varieties in the environment of speaking English in New Zealand (NZ) and North America (US). It further identified students' overall perspectives of pronunciation instruction and acquisition in the inner-circle countries. The study started with a broad question: What are the perceptions of ESL learners in studying pronunciation in inner-circle countries? Then, it was guided by two research questions:

- (1) Is there any difference between ESL learners' expectations of their pronunciation lessons in New Zealand and those in North America?
- (2) Is there any difference between ESL learners' attitudes toward instructors' English accents in New Zealand and those in North America?

## **METHOD**

### **Participants**

The participants were 238 adult ESL students from language institutes in two inner-circle countries. One hundred fifteen were studying in ESL programs in Auckland, NZ and 123 in Arizona and in Georgia, in the US. They came from 14 different language backgrounds; the largest first language groups were Mandarin (n=64), Korean (n=58), Japanese (42). All had at least a high school education, and were placed into language classes according to ESL proficiency tests developed by the language institutes. In the NZ setting, there were 84 females and 31 males aged from 19 to 40. Twenty-five of them volunteered to be interviewed after filling out the open-ended/scalar-response surveys. Twenty-two percent were placed in beginners' level; 53 percent in intermediate, 25 percent in high-intermediate and advanced. In the US setting, there were 77 females and 46 males aged from 17 to 32. Twelve of them participated in both the interviews and the surveys. Twenty-four percent were in beginner's level; 45 percent in intermediate; and 31 percent were high-intermediate and advanced.

### **Procedures**

A survey instrument was designed by adopting an idea in Derwing's (2003) accent questionnaire. Items were also developed on the basis of findings drawn from oral interviews with ESL students as a pilot study. Participants were asked to make scalar judgments on a six-point scale (1=strongly disagree, 6=strongly agree, and NA=not applicable). There was also an open-ended prompt in which the respondents were asked for expectations of their pronunciation lessons and

their attitudes toward ESL instructors in NZ or in the US. The whole survey had 20 questions (10=Likert scale statements and 10=open-ended prompts), a subset of which is reported here (see Appendix). Responses were collected either online or in person. That is, two-thirds of the responses (about 30 responses) in each country were collected online. The interviews were conducted in such a way so as to gain further insights into participants' rationales for providing responses to open-ended questions in the survey. They were recorded either by the researcher or by volunteer participants who also took notes when necessary.

### **Analysis**

The data were examined for difference of ESL learners' expectations and attitudes toward pronunciation instruction and accent varieties between NZ and US settings. The Mann-Whitney U-test was computed to compare the mean of two groups for scalar judgment scores. Interview responses were coded for corresponding open-ended prompts.

### **RESULTS**

In order to answer the research questions, the scalar judgments of respondents were compared through the Mann-Whitney U-test. Initial results revealed that both ESL learners in NZ and the US agreed upon the importance of pronunciation improvement. An overwhelming 93% percent and above reported that pronunciation is important for communication, they are concerned about it, and therefore they want to improve the way they sound very much. [In calculating these results, the study collapsed responses from 1-3 to get a 'disagree' score and 4-6 for the 'agree' score.] In addition, there was no significant difference in participants' scalar judgments on their current pronunciation. ESL students both in NZ and the US reported that good pronunciation made them confident in English, but they themselves believed that they did not have such skills. Over 80 percent of the respondents both in NZ and the US thought that they could recognize the difference between native-like and accented pronunciation in English.

---

*Table 1. Students' Attitudes Towards Pronunciation Studies in NZ and the US (N=238)*

Item descriptor	Group	N	Mean	Sd.	Z	p
Pronunciation is important for communication.	NZ	115	5.10	0.87	-1.78	.100
	US	123	5.36	1.06		
I am concerned about my pronunciation.	NZ	115	5.36	1.06	-1.56	.118
	US	123	5.12	1.08		
I want to improve the way I sound very much	NZ	115	5.23	0.92	-1.31	.191
	US	123	5.34	1.05		
I really want to sound like a native speaker.	NZ	115	4.43	1.48	-3.67	.000
	US	123	5.25	1.14		
Occasionally, I deliberately avoid sounding like a native speaker.	NZ	115	3.48	0.20	-2.34	.019
	US	123	3.02	1.29		
If I have good pronunciation, I will be more confident in English.	NZ	115	5.20	0.91	-1.50	.133
	US	123	5.32	1.04		
I feel that I currently have excellent pronunciation skills.	NZ	115	3.20	0.86	-1.97	.098
	US	123	2.98	1.29		
I believe that my teacher's production provides me with an excellent model of English pronunciation.	NZ	115	3.77	0.19	-7.99	.000
	US	123	5.15	1.03		
It is very confusing to study pronunciation in NZ/US because there are many accents	NZ	115	4.77	1.20	-4.02	.000
	US	123	3.98	1.38		
I can accurately recognize the difference between native-like and nonnative ("accented") pronunciation in English	NZ	115	4.43	0.89	-1.29	.198
	US	123	4.05	1.50		

1=Strongly disagree 2= Disagree 3=Somewhat disagree 4= somewhat agree 5= Agree 6=Strongly agree  
NA=Not applicable

However, in response to the first research question, "is there any difference between ESL learners' expectations of their pronunciation lessons in New Zealand and those in North America?", there were significant differences found among learners' responses between in NZ and in U.S. regarding the following questionnaire statements:

(Item, #4) I really want to sound like a native speaker ( $z = -3.67, p < 000$ )

(Item, #5) Occasionally, I deliberately avoid sounding like a native speaker ( $z = -2.34, p < 05$ ).



Almost 37 percent of respondents did not think it was desirable to sound like a native speaker in NZ whereas only 5 percent disagreed in North America. In addition, 26 percent of the students in NZ indicated that they even occasionally avoided sounding like a native speaker of the target language in contrast with 8 percent of agreement from respondents in the US. Seventeen percent of NZ respondents and 21 percent of US students rejected the idea entirely and marked 'NA (not applicable)' for this question. One Korean female respondent from the interview stated that if she spoke with NZ accent after she went back to Korea, people might laugh at her. A male Thai student who studied English in Auckland for 3 months commented as follows:

(1) I am afraid of having NZ accents... if I stay here long. They don't really sound stupid but very not clear. My friends don't like it anyway.

The answer to the second question concerning the difference between ESL learners' attitudes toward instructors' English accents in New Zealand and those in North America is positive. ESL students in NZ and the US judged the following questionnaire items significantly differently.

(Item, #8) I believe that my teacher's production provides me with an excellent model of English pronunciation ( $z = -7.99, p < .000$ )

(Item, #9) It is very confusing to study pronunciation in NZ/US because there are many accents ( $z = -4.02, p < .000$ )

Forty percent of ESL learners studying English in NZ disagreed that their teachers' production provided them with an excellent model of English pronunciation while 5 percent of respondents in the US disagreed. By the same token, a majority of respondents (87 percent) in NZ agreed that studying pronunciation in NZ was very confusing because there were too many accents. Therefore, they had difficulties studying English in NZ. On the other hand, 13 percent of respondents in the US agreed with this statement. In the opening quotation from the interview, participants in NZ provided support for their responses. Consider the following comments, one from a Taiwanese female student who had studied English in Auckland, NZ for 18 months and the other from a Saudi Arabian female student studying in NZ for 6 months. The third one was from two Korean students who studied English in Auckland for about 4 months.

(2) There are too **many accents** in school. Well ... teachers are from America, England, Australia, India, Asia and, of course here Kiwi. I don't know which sound I have to follow ... once I studied with my American teacher, and then ... with NZ teacher. My Kiwi teacher don't like American accent, I think. ... Sometimes I speak American accent...um... she correct me. I don't know...

(3) Many accents make me confused and even make my English bad.

(4) My teacher said 'vase [va:z]' for 'vase [veiz]'. So I said, teacher, it's a 'vase [veiz]'. Then, the teacher said no it's [va:z]. So I got silent.

---

The following comment in (5) was made by a Turkish student who studied English in the US for 3 months. The comment in (6) was made by a Chinese student who studied English in the US for 6 months.

(5) My teacher is from Russia, I think. I like to speak like her. She has some Russian accent, but it's okay.

(6) I've never thought about my teacher's English. She is American. She has an American accent. It's good and clear. I'm used to her accent, you know.

Evidently, participants in NZ (63 percent) were more dissatisfied with their pronunciation instruction than those in the US (27 percent) when they were asked about their pronunciation lessons (Question 3 in the open-ended questionnaire). Examples of students' responses to this question included 'teacher's confusing models' or 'no specific instruction'. ESL students were further asked an open-ended question: "If you study pronunciation only, the best place would be...". Figure 1 shows that 30 percent of respondents in NZ expressed that they would prefer to go back to their home country if they solely focused on studying pronunciation. America was ranked second with 26.3 percent, England next, and NZ last. On the other hand, 3 percent of ESL students in the US chose their home country, and a majority of them (65 percent) chose U.S.

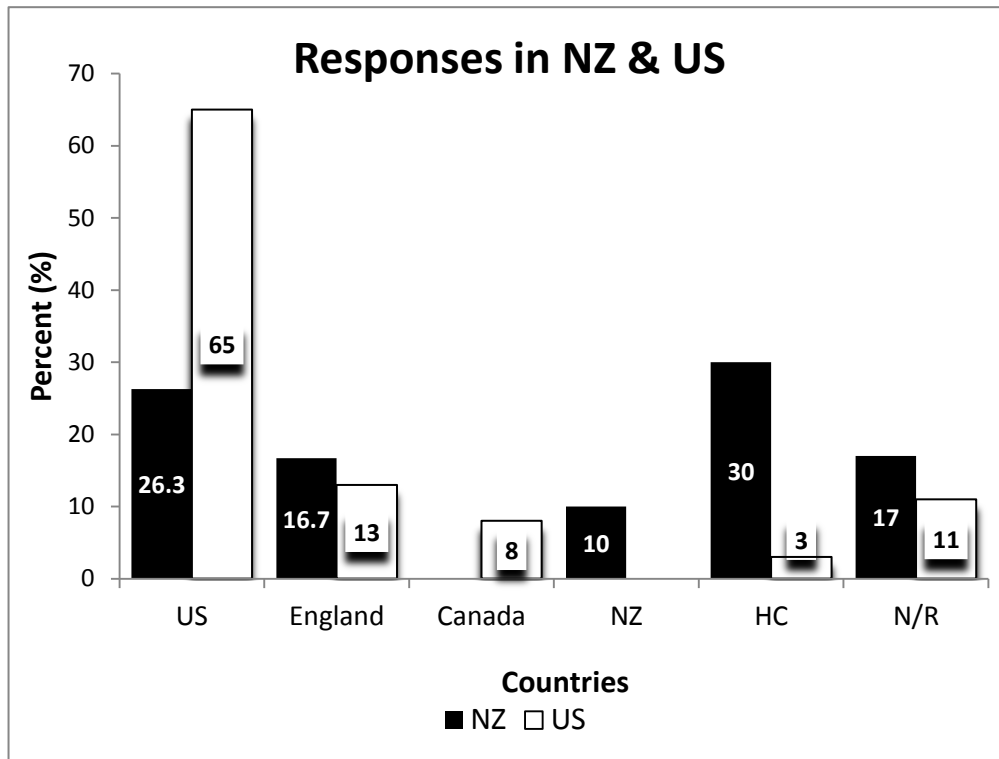


Figure 1. If I just study pronunciation only, the best place would be... (Note. HC=home country; N/R=no response)

Interview comments from a Thai female student in (7) and (8) provide an example of these student attitudes.

(7) My teacher doesn't teach me pronunciation....I have a 'ch' sound problem so badly... but my teachers don't care... I think... just to improve English pronunciation, studying in Thailand would be better than studying in Auckland....

(8) The teacher in my home country understands my pronunciation problems. So they know my difficulty.

## **DISCUSSION**

The study surveyed adult ESL learners who studied English in two inner-circle countries, New Zealand and the United States to examine their expectations of and attitudes toward pronunciation lessons and instructors' accent varieties. Findings revealed that there was no significant difference among learners in the two countries with regard to the need and desire for pronunciation improvement. However, students in NZ were more dissatisfied with their current curriculum of learning pronunciation due to misunderstanding of various models and accents of pronunciation made available to them.

The findings of this study concur with a previous report that ESL learners prefer to model inner-circle standards (Bayard et al, 2002; Timmis, 2002). Literature has often declared that NNSs feel some kind of obligation to acquire 'near-native' English accents in order to be seen – and to see themselves – as successful English speakers, despite the fact that NNSs no longer learn English to communicate primarily with its NSs (Jenkins, 2007, 2009). Not surprisingly, the results of the current research indicated that learners' attachment to inner-circle native speaker models primarily was stronger for North American English. The positive feedback of such US-based students on their current pronunciation models might be formed by students' stereotype of and familiarity with American English (Dalton-Puffer, Kaltenboeck, & Smit, 1997). On the other hand, more than one third (37 percent) of L2 learners in NZ had little desire to sound like native speakers. In addition, one fourth (26 percent) of them even deliberately attempted to avoid sounding like a NZ speaker. Approximately 5-8 percent of students in the US responded in such a manner. Even though these results can be interpreted from a viewpoint of Morley's (1991) learner identity – L2 learners' avoiding a native model to retain their own accent as an indicator of identity – they suggest that learners' preference and attitude toward inner-circle accents vary among the types of models.

From the perspective of World Englishes, mutual intelligibility is a key issue for both listeners and speakers. Nevertheless, learners often have an idealized notion of native-speaker spoken norms, which are particularly related to GA or RP (Timmis, 2002). In this respect, if the inner-circle standards did not meet learners' expectations, L2 learners tended to experience frustration while studying English in a given environment. Forty-percent of L2 learners in NZ reported that teachers did not provide an excellent pronunciation model. Moreover, 87 percent of respondents

---

in NZ said that different varieties of accents confused their study of pronunciation. These NZ results are significantly different from those in U.S. where the native speakers' accent was the preferred model of English for L2 learners. As seen from Taiwanese and Korean student's comment in Excerpts (2) and (4), some ESL teachers might regard their own accent as the most appropriate and consider other varieties incorrect. Perhaps learners' confusion might be caused not by the fact that many varieties of accents were available, but by the fact that there was no comprehensive instruction from pronunciation teachers regarding accent varieties around the world. Consequently, learners built up a negative perspective on poly-models.

In reviewing students' comments in this study, teachers in NZ appeared to have more diverse L1 backgrounds than those in the US (see Excerpt 2). Interestingly, the L1 backgrounds of teachers whose students were surveyed in the US were almost all North American English except for one Russian. Nevertheless, note that students' concern and confusion about their pronunciation models might be also exacerbated by teachers' treatment of accent variation (Excerpts 2 and 4). The general lack of teacher's awareness of World Englishes and accent varieties may be an additional cause of students' pronunciation attitudes. The need for ESL teachers' pronunciation training has been particularly emphasized (Breitkreutz, Derwing, & Rossiter, 2002; Burgess & Spencer, 2000; MacDonald, 2002; Wang & Munro, 2004). Subsequently, teachers' training in current trends in EIL is called for to a great extent.

Students' concern for various models in ESL settings affected their decision on places to study pronunciation. As seen in Figure 1, roughly one third of L2 learners in NZ, as compared to 3 percent of learners in the US, chose their home country for the best place to study English pronunciation. Learners seemed to value teachers whose L1 was the same as theirs [Excerpts (7) and (8)]. This result was surprising because a common belief is that ESL students would improve their target language when they are in an environment where the target language is spoken. According to Excerpts (2) and (4), students seemed to be more confused with inner-circle accent variation than expanding-circle accents. That is, when an inner-circle native speaker, whose accent did not belong to one of the standard models such as GA or RP, taught pronunciation, ESL students appeared to feel much more perplexed. Conversely, learners' attitudes were quite receptive when fluent, non-native English teachers taught pronunciation. [Note that no participant in the US raised the question of dialect in American English in this study.] Overall, the results of the research suggest that students' perceived needs and their expectations require better synchronization in the inner-circle contexts.

## **IMPLICATIONS AND CONCLUSION**

In listening to students' voices through this study, we saw the urgent need for students and teachers to comprehend accent varieties. The review of students' beliefs, attitudes, and perceptions strongly hints that the social relationships between students and teachers are more complex than they may appear in the classroom. Moreover, students' comments imply that ESL teachers' English models have an extremely important effect on the ways in which the learners

relate to learning pronunciation. In other words, teachers' attitudes toward varieties of accents can play a critical role in shaping learners' perspectives and expectations of language learning particularly in pronunciation. Therefore, the role and significance of the teacher's accent need to be further explored.

Furthermore, the current study neither investigated learners' opinions in the setting of speaking RP English, nor examined students' perception of instructor's English accent in the situation of the Outer Circle. It would be very interesting to see how students react to different accent varieties in such environments. Finally, the study only interviewed 25 out of 238 participants. In order to better understand learners' insight into accent varieties, a wide range of in-depth interview would be recommended.

The study started with questions raised regarding (1) which varieties of English to choose for instructional models for pronunciation and (2) whether or not learners' reactions changed if they were exposed to a different variety of inner circle models. The answer to the first question is not to choose a single model for learners considering learners' preferences, but to encourage both teachers and learners to develop a comprehensive understanding of accent varieties in World Englishes. In terms of the issues related to the second question, the findings of this study demonstrated that L2 learners reacted differently as to the inner-circle English models, and that learners' expectations and attitudes toward inner-circle models were very complex and should be carefully taken into consideration in ESL contexts.

The implications of this study concerning the relations among pronunciation training, learner perceptions, and speech variety extend beyond the immediate setting in NZ and US higher education, and in fact pertain to the teaching of World Englishes.

## **ABOUT THE AUTHOR**

**Okim Kang** is an assistant professor of applied linguistics at Northern Arizona University. Her research concerns aspects of second language pronunciation, oral language proficiency assessment, speech perception and language attitudes, and World Englishes. She has investigated prosodic characteristics of nonnative speakers' accented speech and the effect of rater background characteristics on L2 speaking proficiency. Email: [okim.kang@nau.edu](mailto:okim.kang@nau.edu)

---

**REFERENCES**

- Bayard, D. (1990). God help us if we all sound like this: Attitudes to New Zealand and other English accents. In A. Bell and J. Holmes (Eds.), *New Zealand Ways of Speaking English*. (pp. 67-96). Clevedon: Multilingual Matters.
- Bayard, D. (1995). *Kiwitalk: Sociolinguistics and New Zealand society*. Palmerston North: The Dunmore Press Limited.
- Bayard, D., Callois, C., Weatherall, A., & Pittam, J. (2001). Pax American? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics*, 5, 22-49.
- Bayard, D., Gallois, C., Ray, G. B., Weatherall, A., & Sullivan, K. P. H. (2002). *Evaluating English accents worldwide*. Retrieved March 18, 2010, from <http://www.otago.ac.nz/anthropology/Linguistic/Accents.html>
- Bradac, J. J. (1990). Language attitudes and impression formation. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 388-412). Chichester, UK: Wiley.
- Breitkreutz, J. A., Derwing, T.M., & Rossiter, M. J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal*, 19, 51-61.
- Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28, 191-215.
- Crystal, D. (2003). *English as a global language*. Cambridge, England: Cambridge University Press.
- Dalton-Puffer, C., Kaltenboeck, G., & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes*, 16, 115 – 128.
- Derwing, T.M. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review*, 59, 547-566.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39, 379-397.
- Erling, E. J. (2005). The many names of English. *English Today*, 81, 40-44.
- Goodwin, J. (2001). Teaching pronunciation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 117-137). Boston: Heinle & Heinle.

- Jenkins, J. (2000). *The phonology of English as an international language*. Hong Kong: Oxford University Press .
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23, 83–103.
- Jenkins, J. (2003). *World Englishes: A reference book for students*. New York: Routledge.
- Jenkins, J. (2006). Current perspectives on teaching World Englishes and English as a lingua franca. *TESOL Quarterly*, 40, 157–181.
- Jenkins, J. (2007) *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.
- Jenkins, J. (2009). English as a lingua franca: interpretation and attitudes. *World Englishes*, 28, 200-207.
- Kachru, B. B. (1992). *The other tongue: English across cultures*. Urbana: University of Illinois Press.
- Kachru, B. B., and C. L. Nelson. (2001). World Englishes. In A. Burns & C. Coffin (Eds.), *Analyzing English in a Global Context* (pp.9-25). London and New York: The Open University.
- Kang, O., & Rubin, D. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28, 441-456.
- Kirkpatrick, A. (2008). English as the official working language of the Association of Southeast Asian Nations (ASEAN): Features and strategies. *English Today*, 94, 27–34.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. New York: Routledge.
- MacDonald, S. (2002). Pronunciation—views and practices of reluctant teachers. *Prospect*, 17, 3–18.
- Modiano, M. (2001). Linguistic imperialism, cultural integrity, and EIL. *ELT Journal*, 55, 339 – 346.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481-520.
-

- Moyer, A. (2004). *Age, accent and experience in second language acquisition*. Clevedon, England: Multilingual Matters.
- Ryan, E. B., & Carranza, M. (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican-American accented English. *Journal of Personality and Social Psychology*, 31, 855-863.
- Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, 20, 213-223.
- Scales, J., Wennerstrom, A., Richard, D., & Wu., S-H. (2006). Language learners' perceptions of accent. *TESOL Quarterly*, 40, 715-738.
- Timmis, I. (2002). Native-speaker norms and international English: A classroom view. *ELT Journal*, 56, 240 – 249 .
- Wang, J., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32, 539–552.
- Yano, Y. (2001). World Englishes in 2000 and beyond. *World Englishes*, 20, 119 – 131.



## Appendix: Examples of Survey Questions

**Scalar judgment** (1 =Strongly disagree 2= Disagree 3= Somewhat disagree 4= somewhat agree 5= Agree 6= Strongly agree NA=Not applicable)

1. Pronunciation is important for communication.	1	2	3	4	5	6	N/A
2. I am concerned about my pronunciation.	1	2	3	4	5	6	N/A
3. I want to improve the way I sound very much.	1	2	3	4	5	6	N/A
4. I really want to sound like a native speaker	1	2	3	4	5	6	N/A
5 Occasionally, I deliberately avoid sounding like a native speaker.	1	2	3	4	5	6	N/A
6. If I have good pronunciation, I will be more confident in English.	1	2	3	4	5	6	N/A
7. I feel that I currently have excellent pronunciation skills.	1	2	3	4	5	6	N/A
8. I believe that my teacher’s production provides me with an excellent model of English pronunciation.	1	2	3	4	5	6	N/A
9. It is very confusing to study pronunciation in America because there are many accents.	1	2	3	4	5	6	N/A
10. I can accurately recognize the difference between native-like and non-native (“accented”) pronunciation in English.	1	2	3	4	5	6	N/A

### Open-ended questions

1. If I just study pronunciation only, the best place would be (America, my home country and others: \* please specify the place)
2. In order to improve my pronunciation, I want to
  - i. imitate English native speakers or
  - ii. develop my own clear accent .
 \* Please specify the reason.
3. My pronunciation lessons in school are (satisfactory/ unsatisfactory).  
\* Please specify the reason.
4. I came to America to study English because.....
5. The English I prefer is (American, British, Australian, and New Zealand).

Koffi, E. (2010). The pronunciation of <-ED> in coda clusters in Somali-accented English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1<sup>st</sup> Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. (pp. 119-134), Ames, IA: Iowa State University.

---

# **The Pronunciation of <-ED> in Coda Clusters in Somali-Accented English**

**Ettien Koffi**

St. Cloud State University

Clements and Keyser (1985, p. 28) note that the most prevalent syllabic structure found in world languages is the CV pattern, that is, a single Consonant followed by a single Vowel. English far exceeds this minimal requirement by allowing up to four consonants in the coda. This heavy coda structure clashes with the simple Somali CV (C) syllable structure. This paper investigates aspects of the pronunciation difficulties experienced by Somali speakers when the past tense suffix /d/ is added to English verbs whose roots end in CVC. The heavy coda cluster that results from such an affixation leads to frequent instances of phonological interference. Phonological and acoustic data are presented to account for why Somali speakers have a hard time with verbs whose codas ends with /p/, /t/, and /k/.

## **INTRODUCTION**

The goal of this paper is to try and provide answers to the following questions:

When people learn that I know something about the Somali language, the two most frequently asked questions are, ‘Why do they add an [i] to the end of their words?’ and ‘Why do they pronounce all past tense verbs with [tɪd]?’ (Lindsey 2006, p. 62)

I have been asked the same questions more times than I could count. I have also heard Somali speakers of various proficiency levels make the same mistakes in my college classrooms, on television, and in the streets. Being an African and also being a professor of linguistics, people naturally expect me to know the answer to these questions and to provide a remedy for these pronunciation woes. The relentlessness of the questions and the pervasiveness of the errors have finally convinced me “to do something about it.” So, I have added aspects of Somali-accented English phonology to my research agenda. This paper is the first installment of a number of papers that seek to account for Somali English. The present investigation focuses almost exclusively on the pronunciation of the suffix <-ed><sup>1</sup> when it is attached to verbs whose codas consist of only one consonant.

---

## GENERAL BACKGROUND ON THE SOMALI LANGUAGE

Linguists classify Somali as an Afro-Asiatic language. They further sub-classify it as a Cushitic language. The Cushitic family is divided into three main families:<sup>2</sup> North Cushitic, Central Cushitic, and East Cushitic. Somali, along with Afar and Oromo, belongs to the East Cushitic branch. It is generally agreed that there are three major dialects of Somali:<sup>3</sup> the northern dialect, the southern/coastal dialect, and the central dialect. The former is also referred to as Common Somali or Standard Somali. The coastal dialect is known as Benadir, and the central dialect is called “Maay-maay.” It is next to impossible to find out the exact number of Somaliphones in the world. Abdullahi (2000, pp. 20-21) cautiously gives the following figures: 9,500,000 speakers in Somalia proper, 3,500,000 speakers in eastern Ethiopia, 500,000 in northern Kenya, and 250,000 in Djibouti. Almost 20 years ago, there were approximately 14 million Somaliphones in the Horn of Africa. However, because of the civil war that has been raging since the 1980s, there are thousands of Somalis in many parts of the world. By the 1990s, more than 16,000 Somalis lived in the Twin Cities of Minneapolis and Saint-Paul (Abdullahi 2000, p. 22). Current estimates of Somalis in Minnesota vary from 60,000 to 80,000.<sup>4</sup> Admittedly, there are more Somalis in south-central Minnesota than anywhere else in the world except in the Horn of Africa.

### Overview of Somali Consonant Inventory

I will dispense with a balanced treatment of the phonological systems of English and Somali. I favor Somali in this overview because I assume that basic facts about English phonology are known to the reader. However, this is not so for Somali. Consequently, much of the focus in this section will be on elementary aspects of Somali phonology. The bulk of the information comes from Saeed (1999, pp. 7-51).<sup>5</sup>

At first glance, the Somali consonant chart is similar to that of English in many respects. Both languages share a great number of stops, namely /t, k, ʔ, b, d, g/. However, one English consonant that is conspicuously missing in Somali is /p/. More will be said about this sound shortly. Somali has other stops that English lacks, specifically /d, ɠ/. In the fricatives column, the two languages share four consonants /f, s, ʃ, h/. The notable English fricatives that are absent in Somali are /v, θ, ð, ʒ/. The Somali fricatives /χ, h, ʕ/ do not exist in English. English has two affricates /tʃ/ and /dʒ/ but Somali only has the voiceless affricate /tʃ/. The English nasal /ŋ/ is also absent from the Somali inventory. The two languages have the liquids /l, r/. However, the two sounds have different places and manners of articulation in the two languages. The Somali /l/ is a retroflex according to Saeed, whereas /l/ is an alveolar in English. Similarly, /r/ is an alveolar trill in Somali whereas it is an alveopalatal approximant in English.

---

Table 1. Somali Consonant Chart

	Bilabial	labiodental	Alveolar	Palatoalveolar	Retroflex	Velar	Uvular	Pharyngeal	Glottal
<b>Stops</b> Voiceless			t			k			ʔ
<b>Stops</b> Voiced	b		d		ɖ	g	g		
<b>Fricatives</b> Voiceless		f	s	ʃ			χ	ħ	h
<b>Fricatives</b> Voiced								ʕ	
<b>Affricates</b> Voiceless				tʃ					
Nasals	m		n						
Trill			r						
Lateral						l			
Glides	w			j					

Just by looking at the similarities and the differences, proponents of the Contrastive Analysis Hypothesis and the Markedness Theory would predict that there would be a significant amount of negative phonetic transfer. This prediction is actually born out when one listens to Somali English. The case of [p] readily stands out and has been analyzed acoustically by Conway (2008). A contrastive analysis would also make the wrong prediction that since [t] is a voiceless alveolar stop in both languages, it would be positively transferred from Somali to English. Nothing could be further from the truth. Our study will show that Somalis have a hard time pronouncing [t] when it occurs in the coda.

### Overview of Somali Vowel Inventory

Somali vowels form a perfect symmetry along two criteria. For the front vs. back series, Saeed (1999, p. 11) lists five vowels for each category. These 10 vowels are split equally between [+ATR] and [-ATR] (Advanced Tongue Root). Table 2 summarizes the salient features of the vocalic system of the language.

Table 2. Somali Vowel Chart

Height	Front	Back	[+ATR]	[-ATR]
High	[i, ɪ]	[u, ʊ]	[i, u]	[ɪ, ʊ]
Mid	[e, ɛ]	[o, ɔ]	[e, o]	[ɛ, ɔ]
Low	[æ]	[ɑ]	[ɑ]	[æ]

Classifying vowels in any language is a challenge. The same is true for Somali. The exact number of Somali vowels and their classifications is disputed.<sup>6</sup>

Numerous sociolinguistic studies have made note of the fact that vowels are the primary carriers of accents. However, since the pronunciation of vowels is outside of the scope of our inquiry, nothing more will be said about them except to report the following impressionistic findings:

Somali and English share a number of the same vowel phonemes and diphthongs. Because of this, problems with pronunciation will not likely come because a student can't produce the vowel in question (Lindsey 2006, p. 47).

Conway (2008, p. 29) concurs with this assessment by stating the following:

English vowels should not be problematic for Somali ELLs in the way consonants are. The Somali front and back vowel series are more marked than English vowels. The one aspect that might cause an issue for a Somali speaker is learning the English pattern of lengthening a vowel before a voiced final stop or fricative.

### **Preliminary Observations about Somali Voiceless Stops**

There are many phonotactic constraints concerning the distribution of Somali consonants and vowels that are worth investigating. However, the consonants [p, t, k] receive the lion's share of attention in this paper because they cause the most trouble to Somali speakers, especially when they occur in English syllable codas. In fact, it does not take long for a casual listener of Somali-accented English to realize that many speakers have difficulties with these voiceless stops.

The difficulty that Somalis have with the sound [p] comes as a surprise to many speakers of North American English. First, the sound [p] occurs frequently. Whitney (2004) lists it as the 15<sup>th</sup> most frequent sound in English. Secondly, from the point of view of articulatory phonetics, it does not take much effort to close the lower and upper lips and blurt out a <p>. Thirdly, [p] occurs in 89 percent of the languages of the world. So, to untrained ears the difficulty that

Somalis have with [p] is baffling. However, Somali is not alone among African languages in its lack of [p]. In fact, Clements and Rialland (2005, p. 26) report that [p] is missing from the phonemic inventory of 63.2 percent of North and East African languages. The pronunciation of [p] is particularly troublesome when it occurs at the beginning of English words. So, a word such as <pop> [p<sup>h</sup>aḻ] may sound like [ḻaḻ] in the ears of a native speaker of American English while a Somali speaker thinks that he/she is saying [p<sup>h</sup>aḻ]. Conway (2008, p. 58) did a Voice Onset Time (VOT) analysis of Somalis' pronunciation of initial [p] and found that both beginners and intermediate ELL students voice [p]. A reproduction from her thesis (p. 56) provides us with the following VOT scores for [p] and [b] in word-initial, medial, and final positions:

*Table 3. The VOT of [p] and [b] in Somali (from Conway, 2008)*

<b>N0</b>	<b>Segments</b>	<b>Initial</b>	<b>Medial</b>	<b>Final</b>
1.	[p]	34 / 71 ms <sup>7</sup>	50 / 60 ms	72 / 60 ms
2.	[b]	33 / 66 ms	54 / 42 ms	65 / 60 ms

VOT (Voice Onset Timing) analysis is a method used by phoneticians to determine if a particular sound is voiced or voiceless. Voiced sounds are produced when the vocal cords come closer together and the air molecules that pass through the glottis causes them to vibrate. For voiceless sounds, the vocal cords are further apart, and so the air molecules pass through freely without causing any vibration. VOT is the time lag between when the vocal cords start vibrating and the release of any stop consonant. It is calculated in milliseconds. This method allows linguists to determine if a stop consonant is voiced or voiceless. For voiced stops, the time gap between the vibration of the vocal cords and the release of the stop is less than 20 ms. Ladefoged (2006, pp. 146-7) shows that in some languages, for voiced consonants, vocal cords start vibrating much earlier, resulting into a negative VOT, as is the case of Sindhi where the VOT for [d] is -130 ms. For voiceless consonants, the VOT is longer because the vocal cords don't start vibrating until the articulation of the next voiced segment. Baart (2010, p. 91) suggests that when the VOT is around 30 ms or longer, we begin to see the formation of aspiration. If, however, the VOT is around 20 ms, Ladefoged opines that the consonant is most likely an unaspirated stop. Furthermore, Ladefoged (2001, p. 128) notes that a typical aspirated English [p] lasts about 60 ms whereas [b] lasts between 10-15 ms. Table 3 shows that there is not a substantial difference between word-initial [p] and [b] in Somali-accented English as far as VOT is concerned.

As it turns out, Americans' perception of Somali's pronunciation of [p] as [b] is born out acoustically. A Somali [p] does not sound like a [p] to an American ear. Studies reported by

Ferrand (2007, p. 267) indicate that Americans' perception of [p] and [b] is categorical. A sound is perceived as a [p] in a word initial position if its VOT is between 40 to 60 ms. Presumably, even six-month old children born to parents who are native speakers of American English can categorically distinguish [p] from [b] on the basis of VOT values. So, even though Somalis think that they are saying [p], the sound that they are producing is perceived by speakers of American English to be [b]. Saeed (1999, p. 8) helps us understand why. In Somali, /b/ is automatically devoiced to [b̥] when it occurs word-initially. Since Somali does not have the /p/ phoneme, speakers tend to use [b̥] to replace the English [p]. However, to a speaker of American English [b̥] is closer to [b] than it is to [p] because it falls short of the 40 ms threshold. Another aspect of the Somali VOTs in Table 3 needs comments. The segments [p] and [b] are acoustically perceived as different because there is a difference of at least 40 ms between the two. However, when we compare the VOTs of these two segments across levels of proficiency, in all cases, the difference is less than 15 ms. This means that, VOT does not correlate with proficiency. In other words, the speech of a Somali who was born outside of the USA or emigrated after puberty will be accented when it comes to producing [p] and [b], irrespective of the environments in which these two segments are found. This assessment is true even for my Somali students who are doing their master degrees in linguistics.

The next troublesome sound is [t]. It is the second most frequent sound English after [ɪ]. According to Whitley (2004) [t] occurs in 97.5 percent of world languages. It exists both in English and in Somali. However, its distribution is severely limited in Somali; it cannot occur in a syllable-final position. Right away one can anticipate the pronunciation difficulties that Somali speakers face when producing the English [t] in syllable codas. These challenges become most obvious when the suffix <-ed> is attached to a voiceless consonant and it is pronounced as [t].

The third most difficult voiceless stop that Somalis are confronted with when speaking English is [k]. It ranks 9<sup>th</sup> in frequency in English. Worldwide, it is found in 96.5 percent of languages. However, like [t], [k] does not occur in the coda of Somali syllables. This is what Saeed (1999, p. 9) has to say about its distribution:

[k] is a voiceless velar plosive, orthographically *k*. As with the *t/d* pair, it has a more restricted distribution than the corresponding voiced plosive *g*, occurring only at the beginning of syllables. It is always pronounced fortis and with aspiration, and does not weaken intervocalically, e.g. [k<sup>h</sup>] in *kàalay* V `come!`, *túke* N `crow`.

As will be seen right below, these distributional constraints underscore the pronunciation challenges that Somalis face when they attempt to pronounce the past tense suffix <-ed> at the end of words. We have limited our inquiry to verb roots with a single coda so as not to muddy the situation even further. Though our study is limited in this regard, the conclusions apply to verbal roots with two or three coda clusters.

---

## Overview of Somali Syllable Structure

Since Kahn's (1980) ground-breaking work on English syllables, it has become customary to diagram syllables as follows:

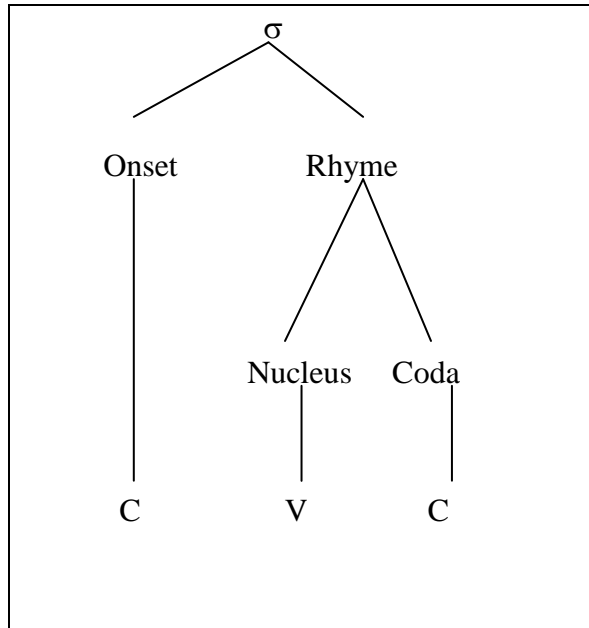


Figure 1. Syllable diagram

Somali has open syllables, that is, syllables that end in a vowel; and closed syllables, those that end with a consonant. Somali has two types of open syllables: CV and CVV syllables. Nothing further will be said about open syllables because they are not of interest to us in the present study. As for closed syllables, there is only one type in Somali, namely CVC. According to Clements and Keyser (1983, p. 28) all languages have the syllable type CV. However, not all languages have syllables that end in a coda. Somali happens to be one of the few African languages that have a CVC syllable structure. When a language has a coda, the resulting syllable can be either heavy or light, (Gordon nd., pp. 2, 5, 27). As for Somali, Orwin (1996, pp. 54, 63) is unsure whether the coda makes the syllable heavy or not. Information gleaned from Saeed (1999, pp. 7-23) seems to suggest that only the following consonants [b], [f], [d], [d̪], [g], [s], [l], [r], [n], [χ] and [h] can occur in the coda. In addition to [t] and [k], the sounds [m] and [tʃ] are not allowed in the coda. Another very important phonotactic constraint that has consequences for the pronunciation of the suffix <-ed> is that Somali does not allow two consonants in the coda.

## The Past Tense Suffix <-ed> and Its Allomorphs

With the background information about the Somali language is in place, let's focus on the suffix <-ed> and its various pronunciations in English. Most phonologists assume that the underlying



phonemic form of the past tense suffix is /d/. Thus, it is customary to state three morphophonological rules for the correct pronunciation of <-ed>. The three allomorphs are [d], [t] and [ɪd]<sup>8</sup> respectively.

The rule for the pronunciation of the inflectional suffix <-ed> is stated as follows:

$$/d/ \rightarrow [t] / [+cons, -voice,] \text{ —————}$$

The rule stipulates that /d/ is pronounced as [t] when it occurs after a voiceless consonant. Thus, if a verb root ends with the consonants [p, f, k, s, ʃ, tʃ], and if the suffix /d/ is added to it, the coda of the past tense verbal stem will be pronounced as follows:

*Table 5. Voiceless Consonants + /d/*

English CVC + /d/	Illustrations
[p]	[pt] <help> → <helped>
[f]	[ft] <surf> → <surfed>
[k]	[kt] <book> → <booked>
[s]	[st] <kiss> → <kissed>
[ʃ]	[ʃt] <fish> → <fished>
[tʃ]	[tʃt] <reach> → <reached>

The second rule for pronouncing the past tense is stated as follows:

$$\emptyset \rightarrow [ə] / [+cons, +alveolar, -nasal, +stop] \text{ —————}$$

This rule states simply that if a verb root ends with the consonants [t] and [d] and the past tense suffix <-ed> is added, the resulting form will be pronounced either as [tɪd] or [dɪd].

Table 6. Alveolar Stop + /d/

English CVC + /d/	Illustrations
$\left. \begin{array}{c} [t] \\ [d] \end{array} \right\} + /d \rightarrow \left\{ \begin{array}{c} [tɪd] \\ [dɪd] \end{array} \right.$	<p>&lt;treat&gt; → &lt;treated&gt;</p> <p>&lt;dread&gt; → &lt;dreaded&gt;</p>

The last rule is usually stated as an elsewhere rule, meaning that if the last consonant is neither of the above, then the past tense suffix <-ed> is pronounced /d/ when the verb ends with voiced consonants, namely, [b, v, g, z, ʒ, dʒ, m, n, l]. (For the purpose of this rule, the semi-vowels /j/ and /w/ behave like consonants when the suffix /d/ is added to verbal roots.) In such cases, the resulting pronunciation rule is as follows:

/d/ → [d] / [+cons, +voice,] \_\_\_\_\_

Table 7. Voiced Consonants + /d/

English CVC + /d/	Illustrations
$\left. \begin{array}{c} [b] \\ [v] \\ [g] \\ [z] \\ [dʒ] \\ [m] \\ [n] \\ [l] \end{array} \right\} + /d \rightarrow \left\{ \begin{array}{c} [bd] \\ [vd] \\ [gd] \\ [zd] \\ [dʒd] \\ [md] \\ [nd] \\ [ld] \end{array} \right.$	<p>&lt;rob&gt; → &lt;robbed&gt;</p> <p>&lt;love&gt; → &lt;loved&gt;</p> <p>&lt;log&gt; → &lt;logged&gt;</p> <p>&lt;buzz&gt; → &lt;buzzed&gt;</p> <p>&lt;judge&gt; → &lt;judged&gt;</p> <p>&lt;comb&gt; → &lt;combed&gt;</p> <p>&lt;tan&gt; → &lt;tanned&gt;</p> <p>&lt;call&gt; → &lt;called&gt;</p>

### Weight Sensitive Explanation of Somali Pronunciation

Lindsey's (2006, p. 62) work has validated my auditory perceptions of Somali pronunciations of <-ed>. She has extensive experience with Somali ELL pronunciation not only as a teacher at the Saint Cloud Technical College in Saint Cloud, Minnesota, but also as someone who has written an MA thesis on the subject. The following sample is gleaned from various examples given in her document:

Table 8. Illustration of Somali Pronunciation of <-ed>

N0	Words	English	Somali
1.	<jumped>	[dʒʌmpt]	[dʒʌmpɪd] <sup>9</sup>
2.	<>walked>	[wakt]	[wakɪd]
3.	<kissed>	[kɪst]	[kɪsɪd]
4.	<treated>	[trɪtɪd]	[trɪtɪd]
5.	<begged>	[bɛgd]	[bɛg]

Of the three allomorphs of <-ed> discussed, Somalis are successful in pronouncing only [ɪd] and [ɪdɪ], as in the words <treated> and <dreaded>. When the inflectional past tense suffix is pronounced [d] or [t] after a consonant, Somali speakers often have trouble pronouncing it accurately. Attempts will be made in the next two sections to explain why.

### Coda Cluster Simplification

The reason Somali speakers cannot produce sequences of Voiceless Consonants + <-ed> or Voiced Consonants + <-ed> has to do with syllable weight. The only attested heavy syllable in Somali is CVV, as in the word in *kàalay* ('come!'). Somali experts are unsure whether Somali CVC syllables should be classified as light or heavy. Even if later studies were to determine that the coda in CVC is heavy, this would not contradict the claim that heavy codas in English are problematic for speakers of Somali. In metrical and autosegmental phonology, a distinction is often made between heavy syllables and super-heavy syllables. The heaviness scale for codas is illustrated as follows: VCCCC > VCCC > VCC > VC (Goldsmith 1990, p. 115). According to Saeed (1999, p. 16) there is not a single word in Somali whose root ends with two consonants. A word such as <dhagxán> (stones) is syllabified as <dhag•xán>. Consequently, English VCC

clusters such as those mentioned in Table 7 are too heavy for Somali ELL speakers who tend to systematically simplify them. Such simplifications rely heavily on two strategies: the deletion of one of the co-occurring consonants or the insertion of a vowel between the two consonants.

***Coda Cluster Simplification by Epenthesis.*** Coda cluster simplification by epenthesis may occur elsewhere in Somali-accented English. However, it is used overwhelmingly when, after adding <-ed> to a verbal root, the resulting coda is pronounced [pt], [kt] or [st], as shown in Table 5. This is particularly true when the speaker is aware that a past tense suffix must be used for past actions or events. In such instances, the vowel [ɪ] is inserted to fulfill the morphophonological requirement of tense. A rule-interaction account for this process may be exemplified by the pronunciation of <jumped> as follows:

Table 9. Derivational Processes

Phonological Processes	English	Somali
1. Underlying Representation	/dʒʌmp/	/dʒʌmb/
2. /p/ Devoicing	NA	dʒʌmb̥
3. Past Tense Affixation	dʒʌmp+d	dʒʌmb̥+d
4. Affix Devoicing Assimilation	dʒʌmpt	NA
5. /ɪ/ Epenthesis	NA	dʒʌmb̥ɪd
6. Phonetic Realization	[dʒʌmpt]	[dʒʌmb̥ɪd]

Such a process of rule interaction may account for why [ɪ] is inserted whenever the past tense suffix is to be realized as [pt], [kt] or [st] in the coda. Or one may ignore this explanation altogether as too speculative and claim that the Somali pronunciation is caused by a spelling-pronunciation interference. But Lindsey (2006, p. 58) refutes the latter explanation:

Some may argue that in both of the above cases of error, students are merely reading what they see. For example, when Somalis encounter the word <kissed> they read it as they see it, thereby producing [kɪsɪd]. This may be a valid point; however, I believe that this is not the case for most Somali speakers. Many English speaking Somali are not literate in English or Somali, and yet they still make these errors. Many of the examples

that I have cited in this section have not come from a reading study but rather from informal conversation with my Somali students. Thus, reading would not be a factor in these instances.

***Coda Cluster Simplification by Deletion.*** Epenthesis is not the only pronunciation strategy that Somali speakers use to simplify two-consonant coda clusters. They frequently resort to deletion in order to avoid weighty codas. More often than not, this happens when the last consonant of a verbal root is voiced, as in Table 7. For instance, Lindsey (2006, p. 54) reports that her students pronounce <begged> as [bɛg] instead of [bɛgd]. This is particularly true of speakers who are not yet strongly aware of the importance of the suffix <-ed> as a marker of the past tense. Conway (2008, p. 65) reaches essentially the same conclusion:

Final /d/ was difficult for both groups to voice [pronounce]. Not one of the beginners was able to voice final /d/. Intermediates also had more difficulty voicing word-final /d/ than those of the other two positions [word-initial and between two vowels]. This is a characteristic that occurred with /b/ as well. It could be that because Americans tend to aspirate word-initial phonemes; Somalis have picked up the emphasis on word-initial phonemes, but not the subtlety of the word-final position.

Cisse-Admison (2009, pp. 68-9) who has experience teaching Somalis at the elementary level and also at the post-secondary level reports that the deletion of the past tense <-ed> is the subject of constant complaints by teachers:

This is a common complaint of teachers of Somali ELLs, especially since grammatical information is often carried by suffixes. Somali ELLs often drop the past tense verb ending *-ed* as in *asked* and the plural *-s* as in *cars* because once the suffix is applied, a consonant cluster is formed in the rhyme of a syllable. In fact, consonant clusters in the rhymes of syllables never occur in Somali ... Thus, for Somali learners of English, a language in which it is possible to have multiple consonant phonemes in a syllable's onset and rhyme, English syllables beginning and ending with multiple consonants can be especially difficult.

There is no consensus on the level of proficiency at which cluster simplification becomes less of an issue for Somali speakers. Lindsey (2006, p. 54) claims that "consonant cluster simplification by deletion strategies are used primarily by beginning students. As students progress to the intermediate level (where most of my students are), they tend to not simply delete their consonant clusters. This has been confirmed by other ELL teachers of Somali who rarely speak of consonant cluster deletion as a serious problem."

---

However, my observations of Somali-accented English show that speakers at various levels of proficiency engage in cluster simplification. Many Somali first year and sophomore students enroll in my courses and I hear some of them simplify their coda clusters. It may be that teachers pay less attention to consonant cluster simplification in the coda in face-to-face interactions because other grammatical features in the discourse environment provide enough redundancy. If a temporal adverb such as “yesterday” occurs in the utterance, the listener is sufficiently situated to ignore the lack of <-ed>. It should be noted right away that final cluster simplification has been attested in the speech of ELL speakers from a variety of linguistic backgrounds. It is also a prevalent feature in pidgins and creoles. Labov (1998, p. 381) advises teachers of AAVE speakers to pay attention to the end of words because codas tend to be simplified in their speech too. Wardhaugh (2007, p. 186) also notes that consonant cluster simplification by deletion occurs in white nonstandard English.

## CONCLUSION

The morpheme <-ed> has been singled out for study because the mispronunciation of inflectional suffixes is frowned upon by native speakers in almost every society. Inflectional morphemes occur with high frequency in all languages. According to Pinker (1999, pp. 124-5), the higher their frequency, the more irregular their form; and yet societies expect their members to acquire these irregular forms and use them accurately. Deviations from established inflectional morphology paradigms are not easily condoned and people who flout these norms may suffer sociolinguistic consequences for their non-conformity. Wardhaugh (2007, pp. 5, 8) explains why:

We will see that there is considerable variation in the speech of any one individual, but there are also definite bounds to that variation: no individual is free to do just exactly what he or she pleases as far as language is concerned. You cannot pronounce words any way you please, inflect or not inflect words such as nouns and verbs arbitrarily, or make drastic alternations in word order in sentences as the mood suits you. [...] Hudson (1996, p. 12) says that we may be impressed by the amount of agreement that is often found among speakers. This agreement goes well beyond what is needed for efficient communication. He particularly points out the conformity we exhibit in using irregular forms, e.g., *went* for the past tense of *go*, *men* as the plural of *man*, and *best* as the superlative of *good*. This irregular morphology is somewhat inefficient; all it shows is our conformity to rules established by others.

The mispronunciation of <-ed> in the coda is a strong marker and has even become stereotypical of Somali-accented English. Wardhaugh (2010, p. 148) defines these terms as follows:

A marker does carry with it social significance. In fact, markers may be potent carriers of social information. People are aware of markers, and the distribution of markers is clearly related to social groupings and to style of speaking. A stereotype is a popular and, therefore conscious characterization of the speech of a particular group.

Armed with the information presented in this paper, ELL teachers of Somali students can now develop pedagogical strategies to improve this aspect of Somali pronunciation. Drills that focus on coda clusters should be part of the pronunciation curriculum. Such a deliberate effort will go a long way toward reducing linguistic prejudice, which as Wardhaugh (2007, p. 117) points out, “is a fact of life, a fact we must recognize.”

## ABOUT THE AUTHOR

Professor Ettien Koffi teaches Interlingual Phonetics and Phonology, Pedagogical Grammar, and Sociolinguistics in the graduate TESOL program at Saint Cloud State University. His areas of specialization and interest are in the application of phonological theory to the elaboration of orthographies for unwritten languages and the acoustic/phonological study of dialect variation. Email: [enkoffi@stcloudstate.edu](mailto:enkoffi@stcloudstate.edu)

## NOTES

---

<sup>1</sup> The following transcription conventions are used throughout the paper: [...] represents actual pronunciation, the convention <...> is for the orthographic form, and /.../ indicates the phonemic (abstract) form.

<sup>2</sup> Abdullahi (2000, pp. 19-20) questions the existence of a South Cushitic family.

<sup>3</sup> Abdullahi (2000, pp. 24-25) notes that there are two schools of thoughts on the number of dialects in Somalia. The Italian School led by Moreno (1955) has identified dozens of dialects whereas the School of Oriental and African Studies led by Andrzejewski (1971) maintains that there are only three dialects.

<sup>4</sup> The information is taken from [www.somalijustice.org](http://www.somalijustice.org) retrieved on 9/10/2009. However, this figure cannot be independently confirmed. A search for an official statement about the number of Somalis in Minnesota has been fruitless. There is no Somali-specific data at the MN Demographic Center. The US Census website does not provide any information either. A Minnesota Public Radio article September 9, 2009 at [http://news.minnesota.publicradio.org/features/200202/04\\_williamsb\\_africans/somalis](http://news.minnesota.publicradio.org/features/200202/04_williamsb_africans/somalis) estimates the number of Somalis in Minnesota to be 15,000.

<sup>5</sup> The consonant chart in Saeed (19, p. 7) omits the places of articulation. This piece of information has been supplied by comparing Saeed's chart with the IPA chart at <http://weston.ruter.net/projects/ipa-chart/view/keyboard/>. Moreover, Saeed does not follow the IPA convention of listing voiceless consonants before listing their voiced counterparts. My adaptation of Saeed's chart complies with the IPA.

<sup>6</sup> The chart of Somali vowels that is presented here is based on Edmondson, Esling, and Harris's acoustic study (n.d, pp. 9-10). I prefer it to Saeed's (1999) classification because there appears to be a number of problems with his classification. For instance on p. 11 he divides the vowels into two series front vs., back, but on p. 12, he presents a chart in which there are three central vowels.

<sup>7</sup> The first number is the average for beginning students and the second the average for intermediate students.

<sup>8</sup> Some native speakers realize it as [æd].

<sup>9</sup> Without the benefit of a VOT analysis, Lindsey may have mistaken the devoiced [b̥] for a [p] and the devoiced [g̥] for a [k].

---

---

**REFERENCES**

- Abdullahi, M. (2000). *Le Somali, dialectes et histoire*. Thèse de Philosophie Doctor (Ph.D.). Faculté des Arts et des Sciences, Département de Linguistique. Université de Montréal, Canada.
- Admison-Cisse, A. (2009). *The effects of pronunciation on Somali spelling errors*. MA Thesis. Saint Cloud State University: St. Cloud, MN.
- Baart, J. (2010). *A field manual of acoustic phonetics*. Dallas, Texas: International Academic Bookstore.
- Clements, G. & Rialland, A. (2005). Africa as a phonological area. Retrieved September 1st, 2009. <http://www.orthographyclearinghouse.org/Paper.pdf>.
- Clements, G. & Keyser, S. (1985). *CV phonology: A generative theory of the syllable*. Cambridge, MA: The MIT Press.
- Conway, E. (2008). *An analysis of Somali pronunciation errors*. M.A Thesis. Hamlin University, Saint Paul, MN.
- Edmondson, J., Esling, J. & Harris, J. (n.d.) *Supraglottal cavity shape, linguistic register, and other phonetic features of Somali*. Retrieved September 1st, 2009. <http://www.ling.uta.edu/~jerry/somali.pdf>.
- Ferrand, C. (2007). *Speech science: An integrated approach to theory and clinical experience, 2<sup>nd</sup> Edition*. New York, NY: Pearson.
- Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Cambridge, MA: Basil Blackwell, Inc.
- Gordon, Matthew. (n.d). *Weight-by-position adjunction and syllable structure*. Retrieved September 1, 2009. <http://www.linguistics/edu/faculty/gordon/weightbypos.pdf>.
- Kahn, D. (1980). *Syllable-based generalizations in English phonology*. New York, NY: Garland Press.
- Ladefoged, P. (2006). *A course in phonetics, 5th edition*. New York, NY: Thomson-Wadsworth.
- Ladefoged, P. (2001). *A course in phonetics, 4<sup>th</sup> edition*. New York, NY: Harcourt College Publishers.



- Labov, W. (1998). Recognizing Black English in the classroom. In D. D. Oaks (Ed.) *Linguistics at Work: A Reader of Applications*, pp. 360-386. New York, NY: Harcourt College Publishers.
- Lindsey, M. (2006). *An ELL teacher's guide to Somali-English interlingual phonology*. MA Thesis. Saint Cloud State University: St. Cloud, MN.
- Orwin, M. (1996). A moraic model of the prosodic phonology of Somali. In *Voice and Power*, ed. Hayward and I.M. Lewis. ALC Supplement 3: 51-71.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York, NY: Basic Books.
- Saeed, J. (1999). *Somali*. Philadelphia: John Benjamins Publishing Company.
- Wardhaugh, R. (2010). *An Introduction to Sociolinguistics, 6<sup>th</sup> edition*. Malden, MA: Blackwell Publishing.
- Wardhaugh, R. (2007). *An introduction to sociolinguistics, 5th Edition*. Malden, MA: Blackwell Publishing.
- Whitley, S. (2004). *The phonicon: A multimedia digital encyclopedia of the sounds of the world's languages*. CD Rom. Dallas, Texas: SIL International.
-

## **Authentic speech and teaching sentence focus**

**Greta Muller Levis**

**John Levis**

Iowa State University

Focus is an essential tool which speakers and listeners use to communicate pragmatic meaning in English regarding the relative importance of information in discourse. We explored using authentic spoken discourse as a source for materials to teach the use of sentence focus illustrating non-final new information in content words. Using examples from academic and non-academic discourse, we found that texts that provided multiple examples of non-final new information were rare. We discuss the issues that arose in our analysis of two texts, including examples where phrases were spoken with multiple focus words, where speakers used focus to mark spoken contrasts, and where anomalous focus placement was hard to describe in terms of either the dominant last content word pattern or in terms of information structure. Finally, we provide recommendations for teaching sentence focus using both authentic and adapted materials.

### **INTRODUCTION**

This paper reports on an attempt to create practice materials for teaching sentence focus using authentic spoken language materials. Specifically, we wanted to identify authentic materials that highlighted focus on non-final new information. Sentence focus is the use of pitch and length to draw the listener's attention to certain words or syllables. Sentence focus has many names in the professional literature. It has been called the tonic (Halliday, 1967), nucleus (Ladd, 1980), sentence stress (Avery & Ehrlich, 1992), highlighting (Bradford, 1988), primary phrase stress (Dickerson, 1989) and a variety of other terms.

In general, focus usually falls on the last content word (N, V, Adj, Adv) of a phrase (thought group or tone unit). Overwhelmingly, these phrases have only one focused syllable. In a study of the prosody of natural speech, Crystal (1969) said: "Less than 10 percent of all nuclei have tails with stressed syllables on the following words. This is a remarkably low proportion, and the generalisation that tonicity falls on the last lexical item is therefore a most reliable one" (p. 224). Translated into the terms used in this study, Crystal is saying that over 90 percent of phrases have focus on the last content word. Less than 10 percent of phrases have focus elsewhere. We will call this the 90/10 rule.

Focus in English has an important pragmatic role, that of marking the information structure of discourse. In other words, overwhelmingly, focus occurs at the ends of phrases, typically on the last content word, but more fundamentally, focus occurs on new information (Halliday, 1967). Discourse organization in English normally presents old information at the beginning of a phrase

---

and pushes new information to the end. Focus is the phonological phenomenon that speakers use to highlight the new information. While focus is a way to mark information structure, it is also the case that speakers mark information structure through grammar and vocabulary. As a result, it is probably better to say that focus usually falls on the last content word (N, V, Adj, Adv) *of new information*.

For teaching purposes, Crystal's claim that 90 percent of phrases have focus on the last content word seems to make the last content word principle an excellent pedagogical rule. However, if we teach that focus falls on the last content word of new information, we should be able to address not only the 90 percent but also much of the remaining 10 percent in which focus is followed by de-emphasized repeated old information. This is illustrated in the constructed example in (1), where focus is likely to be placed on the underlined words. Note that the capitalized words are unlikely to receive focus because they have been previously mentioned. That is, they are old, or given information. They are likely to be de-stressed in discourse.

(1) Today, we'll be discussing weather. . . We'll especially be discussing severe WEATHER. . . things like tornadoes, flooding and hurricanes. . . really SEVERE WEATHER. . . the kind that causes significant and widespread damage to property.

### **Rationale for Teaching Focus**

The importance of teaching sentence focus has been attested by a wide variety of research findings. First, research has shown that, since focus marks the relative importance of information, appropriately placed focus can improve listener comprehension. In addition, it has been found that sentence focus is both learnable and teachable. Finally, researchers believe that focus is important for both ESL and EIL communication.

L. Hahn (2004) found that appropriately placed focus improved listening comprehension. She studied how listeners recalled information from a short lecture. The same lecture was recorded by a bilingual Korean-English speaker with three conditions: correctly placed focus, incorrectly placed focus, and no focus at all (as would happen in Korean, which does not use focus to mark informational prominence). Hahn found that NS listeners who heard the text with the correct (English) prominence placement recalled significantly more information than the listeners who heard the text in either of the other conditions.

Sentence focus also appears to be very learnable. Pennington and Ellis (2000) studied the ability of Cantonese speaking learners of English to remember contrasting utterances containing four different intonational features:

- final pitch movement on tag questions, (rising vs. falling)
  - contrastive focus (Is he driving the BUS? vs. Is HE driving the bus?),
-

- the way pitch marks phrase ends (The fight is over Fred vs. The fight is over, Fred),
- the use of pitch to mark internal structure (lighthouse keeper vs. light housekeeper)

In the first experiment (the untutored condition), the learners' recall of sentences ignored the meaning of all four of the prosodic features. "These advanced Cantonese L1 speakers were good at recognizing previously heard sentences and at rejecting entirely new ones. However, they were poor at rejecting sentences having the same lexis as in previously heard sentences but spoken with different intonation" (p. 380). For example, if they heard the sentence – *He's a good boy, isn't he?* with falling pitch on the tag, and then later heard *He's a good boy, isn't he?* with rising pitch on the tag, subjects in the untutored condition considered these sentences to be the same.

In the second experiment (the tutored condition), subjects were taught how each of the four prosodic features affected meaning and then were given the same test with different sentences. For example, researchers wanted to know, if, in the tutored condition, they heard the sentence - *Is he driving the BUS?* and then later heard -*Is HE driving the bus?* with contrastive focus on *HE*, were the subjects able to identify the second sentence as different because the prosody was different? Only one of the four features showed an effect of instruction, contrastive focus. The ability of the subjects to differentiate based on the other prosodic cues was unchanged, suggesting that instruction may be particularly effective with focus.

In another study, M. K. Hahn (2002) found that it was effective to teach specific rules for placement of sentence focus. Hahn examined performance on a test of pronunciation at 3 times: before instruction (T1), after a semester of instruction (T2) and long after the semester was over (T3). T3 varied from 1-4 years, depending on whether the participants were still at the university where instruction took place. All focus rules showed evidence of learning at T2, but only some of the focus rules showed continued learning at T3. Other focus rules showed significant backsliding at T3, suggesting that some rules were not fully internalized. However, enough of the rules showed significant progress at T3 to indicate that improvement in focus is possible through predictive rule teaching.

Finally, focus is a suprasegmental that appears to be important for both English as a Second Language (ESL) and English as a Lingua Franca (ELF) contexts (L. Hahn, 2004; Jenkins, 2000). It is clear that sentence focus plays a significant role in ESL contexts, where native speaking interlocutors will use and expect others to use focus to mark information structure. Jenkins also argues that focus is critical for ELF communication, in which non-native speakers speak to other non-native speakers. Indeed, focus is the only prosodic feature argued to be essential in her Lingua Franca Core.

## GOALS OF STUDY

The goals of our study were modest and pedagogical. We wanted a variety of discourse level passages that we could use with our students to practice both the placement of focus on new information and the de-stressing of old information. We had previously used a passage from Linda Grant's pronunciation book *Well Said* (1993) for this kind of practice. The passage, on the topic of pollution, has been pedagogically effective at demonstrating focus on non-final new information and includes de-stressed old information in final content words. The passage is reproduced in (2). The placement of focus is underlined. Note how following old information, which is capitalized, is de-stressed after its first mention even when it is found in content words at the end of a phrase.

(2)

OK, today we'll continue our discussion of pollution. Yesterday we defined POLLUTION. Today we'll talk about the impact of POLLUTION...its far-reaching effects. Many people think pollution is just a problem for scientists. But it's not JUST A PROBLEM FOR SCIENTISTS. It's a problem that affects everyone. Since it affects human lives, it's a health PROBLEM. Since it affects property, it's an economic PROBLEM. And since it affects our appreciation of nature, it's an aesthetic PROBLEM. (p. 119).

The passage itself has been extremely useful in our teaching, but it was the only such passage we had, and we felt the need for more practice materials to help our students extend what they had learned to new contexts, and hopefully, to their own speech. We also wanted to create materials that were based on authentic language, and hopefully, to use authentic speech that reflected the principles exemplified in the "Pollution" passage.

### Texts Used for Analysis

To do this, we identified a source likely to give us similar language to that in the "Pollution" passage: online classroom lectures and public lectures from University of California, Berkeley classes and from University of California Television (both hosted by YouTube). We watched a large number of such lectures to identify potential passages. Finding passages that were loaded with examples of non-final new information was not easy, but we did find several passages that were promising. The passage we will present in this paper is from a lecture by Burney LeBoeuf, called "Mother Nature with Seals: Revelations from Long-Term Study." (<http://www.youtube.com/watch?v=0bo--3B4NGc>). The section we discuss in this paper is from 0:37-1:07 of the recording, effectively at the beginning of the lecture. This means that the section is unlikely to make reference to a previous discourse context that may affect focus placement.

In addition, some colleagues suggested that academic speech may have issues not present in less scripted environments, and that we should also look at a different genre of spoken language. Ultimately, we ended up examining a Comedy Central interview between Jon Stewart and Jim

---

Cramer, on *The Daily Show*, March 12 2009, Pt. 2.

(<http://www.thedailyshow.com/video/index.jhtml?videoId=220538&title=jim-cramer-pt.-2&byDate=true>). The sections we will discuss occur between 2:44-4:19.

The Stewart-Cramer interview was a bit of a sensation when it aired and soon went viral. The financial system of the US appeared to be in freefall at the time, and there was a desire to find and expose people who had created, or at least had a vested interest, in the practices that had led to the crisis. Jim Cramer was the host of a CNBC television program called Mad Money, and had advocated many of the practices that appeared to contribute to the financial crisis. Jon Stewart, the host of the Daily Show, is ostensibly a comedian, but has the reputation of being an astute political commentator. The interview that occurs is quite confrontational although it also appears to have a darkly comedic tone as well.

## RESULTS

### Academic Lecture - LeBoeuf

On first listening, the LeBoeuf lecture included just the kind of thing we were looking for. The transcript is reproduced in (3) with the expected focus words underlined and phrases numbered for convenience. It includes final content words and phrases with content words which are old information and thus should be de-stressed (phrases 2, 9, 10).

(3)

1. Long-term studies... 2. uh ecological studies... 3. are not simply advantageous for addressing certain behavioral questions... 4. they're absolutely essential... 5. for understanding how an animal copes with life... 6. understanding the factors that shape its evolution... 7. especially the environmental ones... 8. and why the animal actually looks the way it does... 9. and behaves the way it does... 10. you can't answer those questions without a long-term study... 11. This is vital.

Unfortunately, what sounded good at first listening actually included a number of things that didn't fit the "last content word in the new information" pattern. This is reproduced in (4), with actual focus words in *italics*. If a word is underlined and italicized, then focus was placed on the expected word according to the last content word pattern. If a word is in italics but not underlined, placement was unexpected according to the rule, as in phrases 3, 5, 6, 8, and 10.

(4)

1. Long-term *studies*... 2. uh *ecological* studies... 3. are *not* simply advantageous for addressing certain *behavioral* questions... 4. they're absolutely *essential*... 5. for understanding how an animal *cope*s with life... 6. understanding the factors that *shape* its evolution... 7. especially the

*environmental* ones... 8. and why the animal actually *looks* the way it *does*... 9. And *behaves* the way it does... 10. you can't *answer* those questions *without* a long-term study... 11. this is *vital*.

There are two patterns to these anomalies. The first pattern we see is that there may be more than one focus in a phrase, as in phrases 6, 8, and 10. The expected word was in focus, but in each phrase another word was also in focus. This is not unusual, actually, and our pedagogical rule that there only be one focus per thought group is likely at fault. Bolinger (1986) called this the hat pattern because the early and late focuses are like two peaks of a man's hat with a slight dip in between. It may even be possible to ignore the early focus word in teaching, but there will be students who will hear the early focus and ask about it. If we wish to re-record the passage in our own voice, there will be no problem. However, if we wish to use authentic materials with the voice of the original speaker, we must recognize this anomaly.

A second issue occurred where focus is placed on a word that is not the last content word of new information as in phrases 3 and 5 in the LeBoeuf lecture, reproduced in (5). In phrase 3, *questions* is de-stressed even though it has not been mentioned. The same is true for the backgrounded pronunciation of *life* in phrase 5.

(5)

3. addressing certain *behavioral questions*... 4. they're absolutely *essential*... 5. for understanding how an animal *cope*s with *life*

It is possible that the speaker was de-stressing words that were assumed in his own knowledge of the topic, but this does not make for a good pedagogical rule. Bolinger (1972), in discussing this tendency of speaker to accent unexpected words, wrote an article called "Accent is predictable -- if you're a mind reader." Perhaps more troubling from the point of view of providing a rule is just how natural and normal the speaker sounds. There is nothing odd about his focus placement, but explaining his decisions to a nonnative speaker is unlikely to be easy. This is another problem with using authentic speech as teaching materials for focus.

### **Non-Academic Text - Stewart and Cramer**

The non-academic passage was similar in that it included multiple focuses in one phrase, even including sections in which Jon Stewart put focus on almost every word as he punched home his argument to Jim Cramer, as in (6), where there are 10 italicized focus words in six phrases.

(6)

1. *CNBC*... 2. *could be*... 3. an *incredibly*... 4. *powerful tool of illumination*... 5. for people that *believe* that... 6. there are *two markets*

---

Another anomaly showed up in the Daily Show interview. This involved the use of shell nouns, that is, category nouns that have no real content outside of naming a category for spoken classification. In other words, they are nouns that act like pronouns. In excerpt (7), Jim Cramer uses the word *thing* as a shell noun in phrases 1 and 7, each time with different referents, and *thing* does not receive focus. In other parts of the Daily Show text, nouns such as *stuff*, *guys*, and *these guys* all behaved similarly.

(7)

1. But my *second thing* is... 2. is I talk about the *shorts* every single *night*.... 3. I've *got*... 4. people in *Congress*... 5. who I've been *working* with... 6. to try to get the *uptick rule*.... 7. It's a *technical thing*...

A fourth issue is that there may be constructions or idioms that have their own focal pattern. In (7) the compound construction of *uptick rule* in phrase 6 governs where focus is placed. Likewise, in (8), we see *feels like* with the focus on *feel* rather than *like*, the last content word of phrases 1 and 3. Again, this is normal. Putting focus on *like* would sound very odd. We do not know how common these kinds of intonational idioms are, but it is likely that they are not unusual in authentic language.

(8)

1. So what it *feels like* to us... 2. and I'm speaking *purely* as a *layman*... 3. it *feels* like... 4. we are *capitalizing*... 5. *your adventure*... 6. by *our pension*...

The final issue is that it was not unusual for speakers to use contrasts, and even multiple contrasts, within thought groups, as in (8). In this excerpt, Jon Stewart uses focus on pronominal forms (which are not content words) in phrases 1, 4, 5, and 6 to call attention to the contrast between average people and the financial sector people he accuses of acting unethically. In addition, we see the multiple contrasts in phrases 5 and 6 (*your adventure* and *our pension*). These uses of focus, while perfectly normal, do not follow the 90/10 rule nor do they have focus on content words. They make sense, but they are not the kind of authentic material that can be used without adaptation or explanation.

## DISCUSSION

Our original goal was to use authentic discourse to create materials for teaching sentence focus. We especially were interested in texts similar to the “Pollution” text that fronted the use of focus for non-final new information.

The first thing we found was that most texts had no more than 1-2 examples of non-final new information. This should not be surprising because of the 90/10 rule. The texts that we discuss in



this paper were unusual in that they had a greater number of examples of non-final new information.

Secondly, it was actually difficult to find old information in content words following the focus. Most such old information was in pro-forms, for example, *especially the environmental ones* (LeBoeuf). Again, this follows the 90/10 rule and should not have surprised us. While these examples are valuable for practicing de-stressing, they did not provide the kind of practice we were seeking.

Our third discovery was that phrases with multiple focus points are not unusual in authentic discourse. Speakers do this to convey special importance to more than one part of the text, but it can be confusing to learners. For introducing focus, it is useful to ask students to pay attention to the dominant pattern, one focus per phrase, usually final and non-final focus only in predictable positions. This is often challenging enough. Authentic discourse that has multiple focus words in a phrase does not do a good job of exemplifying the dominant pattern of one focus word per phrase and can give learners a false sense of freedom to put focus on any word they like. Of course, the authentic discourse can be re-recorded in the teacher's voice, but it seems better to use authentic language authentically if at all possible.

Fourth, we found that some of the 10 percent non-final focuses exemplified pronunciation points that would need to be taught separately. This includes lexical units like compound nouns (*the uptick rule*) and intonational idioms (*feels like*). It also includes the separate but crucial category of contrasts, since the emphasis of contrasts seems to trump every other rule, including normal word stress or normal de-stressing of pro-forms (we are capitalizing your adventure by our pension).

Finally, we found that any authentic text that seems suitable often turns out to include something weird. This may be something as straightforward as obscure content or difficult vocabulary, or reference to content outside the text. Or it may include some unusual focus that makes perfect sense to the speaker, but that does not seem to fit known pedagogical rules in any easily explainable way.

## **Recommendations**

We started out looking for authentic texts to extend our students' learning of the principle of focus being used to mark non-final new information and the de-stressing of following old information, as exemplified in the "Pollution" text. We found that such texts were not common, largely because speakers naturally push old information into pro-forms or delete it entirely. In other words, speakers typically follow the principle of placing focus on new information, but focus on new information still may look like the last content word rule.

---

When we consider the more fundamental role of focus in calling attention to new information, it becomes clear that we need to teach students not only how to pronounce focus but also how to order their information to communicate effectively. This suggests that if learners are to understand new information placement, we need to move beyond pronunciation issues alone and describe how information is structured, including practice in placing old information early in the utterance, deleting old information, or putting it into pro-forms.

The pronunciation issues remain critically important, however. We recommend that instruction on focus put sufficient emphasis on the last content word rule, the 90 percent of all utterances, including plenty of opportunity for de-stressing final pro-forms. The authentic texts we examined show many examples of this, and these kinds of texts are severely underrepresented in pronunciation teaching materials.

Finally, explicit teaching of the remaining ten percent, where focus does not fall on the last content word needs to be addressed. Some elements, such as construction stress and intonational idioms, should be taught separately; it should then be explained how they fit into the dominant rule. Others, such as contrasts or the need for explicit repetition of old information in lectures, need special consideration. They should be taught not only from a pronunciation point of view but also a discourse meaning point of view. Although it may appear that the last content word rule is enough, such a rule does not teach learners to create their own meaning or to clearly understand how others create meaning in speech.

We were trying to find authentic texts that would clearly illustrate the underlying principle of focus marking information structure so that our students could see the patterns clearly. The goal of such predictive work is to make our students autonomous users of English so that they can understand others and communicate their meanings clearly. We found that such texts do exist, but are probably better used in adapted form. The system of focus in English follows patterns, especially that of information structure and contrast, but it is also the case that speakers may and do use focus to call attention to information that may not obviously follow the dominant patterns.

## **ABOUT THE AUTHORS**

Greta M. Levis is a senior lecturer in TESL and Applied Linguistics at Iowa State University. She has over 20 years of experience in International Teaching Assistant training at the University of Illinois, North Carolina State University and Iowa State. Her expertise includes the development of teaching skills and English pronunciation. She teaches courses in ITA training, composition, and world languages. She is the co-author of a forthcoming pronunciation textbook, *Pronunciation for a Purpose* (University of Michigan Press). Email: [gmlevis@iastate.edu](mailto:gmlevis@iastate.edu)

John M. Levis is Associate Professor of TESL and Applied Linguistics at Iowa State University where he teaches courses in TESOL methods and materials, sociolinguistics, general linguistics, dialects and the teaching of pronunciation. His research interests are in the teaching of pronunciation, English intonation, and speech intelligibility. He has published in a variety of professional journals such as *TESOL Quarterly*, *Applied Linguistics*, *ELT Journal*, *System*, *TESOL Journal* and *World Englishes*. He is the co-author of a forthcoming pronunciation textbook, *Pronunciation for a Purpose* (University of Michigan Press). Email: [jlevis@iastate.edu](mailto:jlevis@iastate.edu)

## REFERENCES

- Avery, P. & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.
- Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, 48(3), 633-644.
- Bolinger, D. (1986). *Intonation and its uses*. Stanford: Stanford University Press.
- Bradford, B. (1988). *Intonation in context*. Cambridge: Cambridge University Press.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- Dickerson, W. (1989). *Stress in the speech stream*. Urbana: University of Illinois Press.
- Grant, L. (1993). *Well said*. Boston: Heinle & Heinle.
- Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201-223.
- Hahn, M. K. (2002). *The persistence of learned primary phrase stress patterns among learners of English*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Halliday, M. (1967). *Intonation and grammar in British English*. The Hague: Mouton.
- Jenkins, J. (2000). *The pronunciation of English as an international language*. Oxford: Oxford University Press.
- Ladd, D. R. (1980). *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.
- Pennington, M. & Ellis, N. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, 84(3), 372-389.
-