

## BENEFITS OF VISUAL FEEDBACK ON SEGMENTAL PRODUCTION IN THE L2 CLASSROOM

**Daniel J. Olson, Purdue University**

While a growing body of research has established the benefits of pronunciation training on second language (L2) production, these benefits have yet to be incorporated into the general skills language classroom in a systematic manner. Furthermore, although relatively new speech analysis software has been shown to be useful in providing visual feedback for L2 suprasegmental (i.e., intonation) production, there is a relative lack of research on its potential implementation for segmental instruction. The current paper presents a systematic analysis of the effectiveness of a visual feedback paradigm (VFP), in an L2 Spanish classroom, as a pedagogical method for pronunciation teaching at the segmental level (i.e., Spanish intervocalic stops). Results demonstrate a significant improvement of L2 stop production relative to a control group receiving traditional pronunciation feedback. Furthermore, findings demonstrate that VFP leads to small incremental gains. Discussion addresses the role of VFP on segmental production and the potential practical implementations of visual feedback in the lower-level, general skills language classroom.

**Keywords:** Pronunciation, Computer-Assisted Language Learning, Second Language Learning, Phonetics

**APA Citation:** Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology*, 18(3), 173–192. Retrieved from <http://llt.msu.edu/issues/october2014/olson.pdf>

**Received:** November 18, 2013; **Accepted:** February 27, 2014; **Published:** October 1, 2014

**Copyright:** © Daniel J. Olson

### INTRODUCTION

Within second language acquisition and applied linguistics research, the instruction of second language (L2) pronunciation has often taken a back seat to the more prominent aspects of syntax and morphology (Deng et al., 2009). This relative secondary role has led some researchers to claim that the study and teaching of pronunciation has been “marginalized” (Derwing & Munro, 2005, p. 382). As a growing body of research has steadily built a case for the benefits of pronunciation instruction, there remains a disconnect between the laboratory-based empirical research and pedagogical classroom implementations (e.g., Wang & Munro, 2004), a general lack of pedagogical materials for the teaching of pronunciation, and high levels of instructor uncertainty (e.g., Foote, Holtby, & Derwing, 2011). The lack of empirically proven classroom methods is most acute at the lower-levels of language instruction, as materials focused on pronunciation are often geared towards advanced learners of a second language. Furthermore, with the advent of easily accessible speech analysis technology, researchers and instructors are faced with questions about if and how such technology may be applicable in the L2 general skills classroom.

In light of these gaps in both research and implementation of methods for pronunciation instruction in the L2 classroom, this paper has two main goals. First, drawing on emerging research on the use of speech analysis technology for phonetic segmental instruction, this paper seeks to provide empirical evidence for the effectiveness of a visual feedback paradigm (VFP) on L2 segmental production in a classroom setting. As a second, but closely related goal, this paper investigates the generalizability of pronunciation gains, to determine if production improvements resulting from a VFP are transferrable to phonetic phenomena not directly addressed in the pedagogical activities.

## LITERATURE REVIEW

### Pronunciation Instruction in the L2 Classroom

For researchers of pronunciation instruction, a topic of primary interest has been whether or not training has any significant effect on L2 pronunciation, and if so to what extent. To that end, a number of studies have worked to isolate the effect of various types of training, with pre- and post-test evaluations (e.g., Archibald, 1998; Derwing, Munro, & Wiebe, 1998; Derwing & Rossiter, 2003; Elliot, 1995; Simões, 1996; Zampini, 1996). This line of research has examined a variety of L2 phonetic features, including vowel quality, rhotics and trills, voice onset time, liquids, and prosodic features, and has regularly found positive effects of training, albeit to widely varying degrees. Furthermore, these studies have shown pronunciation improvements resulting from a wide variety of treatments, including auditory discrimination training (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997), explicit articulatory instruction (Castino, 1996), awareness training (Pennington & Ellis, 2000), high variability phonetic training (Thomson, 2012) and mixed-approaches (Couper, 2003; Derwing, Munro, & Wiebe, 1997; Elliot, 1999; González-Bueno, 1997; Lord, 2005, 2010). While such findings have been subject to some variation (e.g., Suter, 1976), there is general consensus in the literature that pronunciation training has been shown to improve pronunciation.

While L2 pronunciation has been the subject of an active research agenda, it is clear that the benefits of such investigation have not been systematically incorporated into pedagogical approaches to language teaching (Levis & Grant, 2003), as evidenced by the lack of adequate materials, the largely ad-hoc pronunciation methods such as corrective feedback (Olson, 2014), and instructor uncertainty regarding pronunciation instruction (Foote, Holtby, & Derwing, 2011). The lack of pronunciation materials is most acute at the lower-levels of language learning, with most materials having been designed for advanced learners or for stand-alone pronunciation courses (Darcy, Ewert, & Lidster, 2012). When pronunciation is included in general skills materials, it often lacks sufficient detail, contains inaccurate or inconsistent information, and/or is given minimal attention (for Spanish, see Arteaga, 2000; for French, see Miller, 2012; for ESL, see Derwing, Diepenbroek, & Foote, 2012). With respect to the current methods used for pronunciation instruction in the general skills classroom, a recent survey in the lower-level university L2 classroom found that most instructors included pronunciation in an impromptu manner, with modeling or corrective feedback as the most frequently employed method (Olson, 2014). Furthermore, surveyed instructors spent an average of less than 3 minutes per class period on pronunciation, considered to be insufficient by a majority of the instructors polled (77%). These results parallel other recent findings through both observational methods (Foote, Trofimovich, Collins, & Soler Urzúa, 2013) and instructor self-reports (Foote, Holtby, & Derwing, 2011; Murphy, 2011), demonstrating that pronunciation training is dedicated relatively little class time and consists most commonly of corrective feedback. Finally, instructors have been shown to lack both formal training and confidence to teach pronunciation (Burgess & Spencer, 2000; Derwing, 2010), consistently seeking professional development in the area (Breitkreutz, Derwing, & Rossiter, 2002; Burns, 2006; MacDonald, 2002).

In addition to a general lack of pedagogical materials and methods for pronunciation training, many researchers have noted the lack of investigation dedicated to the teaching of phonetics relative to other language domains, namely syntax and morphology (Deng et al., 2009; Derwing & Munro, 2005; Leather, 2000; Levis, 1999; Lord, 2005; Major, 1998). As such, while a number of studies have demonstrated a positive effect of pronunciation instruction for L2 learners, there appears to be inadequate pedagogical materials and a relative lack of classroom-based empirical research. These shortcomings have led to a “marginalization” of pronunciation in the classroom.

### Visual Feedback for L2 Pronunciation Instruction

Significant advances in technology in the last 40 years have led to the creation of a number of speech analysis tools. While early tools were limited in scope and accessibility, more recent developments are

powerful, freely accessible, and extremely versatile. A growing number of researchers have sought to determine if these tools might serve to improve L2 pronunciation, and if so, how they might be incorporated into the language classroom.

While such technology has come in a variety of forms, one of the most fruitful avenues for research has been the visual feedback paradigm (VFP), also termed electronic visual feedback, as well as the earlier iteration of model visual input. With the advent of software such as *Visi-Pitch* (Kay Elemetrics, 1986), researchers have investigated the benefits of visual analysis of self-produced speech by non-native speakers and/or visual modeling of native speaker productions (de Bot, 1983). Although early work focused more on native models rather than on feedback, the current typical VFP consists of: (a) a non-native speaker recording the stimuli; (b) a visual display of the speech feature, most commonly the intonation contour; (c) a visual display of a native speaker production for comparison, often accompanied by a corresponding auditory presentation; and (d) a re-recording on the part of the non-native participant attempting to match the native-speaker productions (although for transient visual feedback, see Hincks & Edlund, 2009).

Within this framework, suprasegmental speech features, in particular intonation contours, have been among the most studied (Anderson-Hsieh, 1992, 1994; Chun, 1989, 1998, 2002; Molholt, 1988). Much of the early work was conducted by de Bot and colleagues, and showed potential improvements in L2 learner production through visual presentation of native speaker-produced model intonation contours (de Bot, 1980; de Bot & Mailfert, 1982; Weltens & de Bot, 1984). The gains made by L2 learners using a combination of visual and auditory input, were found to be superior to the gains made solely with auditory feedback (de Bot, 1983). More recently, Levis and Pickering (2004) have noted that such a paradigm can be useful for instruction of both local and discourse level pitch contours. With respect to the nature of gains for suprasegmental features, Hardison (2004) showed that improvements in intonation contour productions are generalizable to novel utterances, not included in the initial training (for caution on generalizability, see Seferoğlu, 2005).

Prosodic features represented an ideal starting point for investigation of VFP for a number of reasons, including both practical and pedagogical. Specifically, early speech software was able to produce accurate representations of intonation contours, and interpretation of the contours was found to be relatively intuitive for speakers (Léon & Martin, 1972). While early studies focused specifically on pitch contour, Chun (2007) makes a call to extend the focus beyond the suprasegmental level. Further technological developments and increased accessibility have allowed for investigation of the effects of VFPs on individual speech segments, such as syllables, consonants, and vowels. For example, Lambacher (1999) detailed a VFP implemented in a second language classroom, focused on segmental features (see also Molholt, 1988). Using spectrograms, the VFP provided learners with visual representations of a variety of speech sounds: stops, fricatives, and liquids. While Lambacher (1999) sought only to provide a possible method, subsequent work has aimed to evaluate the quantifiable benefits of a VFP on segmental features, with mixed results. Saito (2007), using PRAAT speech analysis software, for example, showed an increase in vowel pronunciation accuracy for Japanese learners of English who were presented with visual images of their own productions to compare to those of native-English speakers. Similarly, positive results have been found for waveform displays, in conjunction with perceptual training, in the acquisition of two different length contrasts, including vowel length (Okuno, 2013), and singleton/geminate contrasts (Motohashi-Saigo & Hardison, 2009). For example, Motohashi-Saigo and Hardison (2009) reported significantly improved geminate production accuracy in L2 English learners of Japanese, as rated by native listeners, following an audio-visual perceptual training involving waveform displays. In contrast, Ruellot (2011) investigated the benefits of a VFP on vowel production (i.e., /u/ and /y/) in native-English L2 learners of French, and found no overall improvement. Correspondingly, Carey (2004) investigated the benefits of a VFP on vowel production, using *Kay Sona-Match* software, with varying results for different vowels. Thus, while the bulk of the work has examined the effect of visual feedback on

suprasegmental production, there is both clear promise and a need for further investigation of the utility of VFP on segmental production.

It is important to note that much of the previous research has implemented visual feedback within the laboratory (for intonation, see de Bot, 1983; for segmental features, see Saito, 2007), or as part of a larger phonetic-focused course for advanced learners (Lord, 2010). While as a whole, this research demonstrated significant gains in L2 pronunciation, it remains to be seen if such benefits may also be found through implementation of VFP into the lower-level L2 classroom setting where students are still developing their general skills.

### **Concerns about VFP and Speech Analysis Software**

It is worth noting that a number of researchers have raised concerns over the use of stock speech analysis programs, particularly in the L2 classroom. These concerns have predominantly been about practical issues, rather than theoretical ones. Some authors have claimed that speech analysis software was designed for researchers, and is thus ill-suited to the needs of the student (Derwing, 2010; Setter & Jenkins, 2005; Wang & Munro, 2004). In addition, claims can be made that interpretation of visual presentations may not be intuitive for non-trained students (e.g., Chun, 1998), although some forms of visual feedback, such as intonation contours (e.g., Hardison, 2004) may be easier to interpret than others (e.g., spectrograms). Finally, given that much of the previous research on the effect of the VFP has taken place in a laboratory setting, not within the actual L2 classroom, there may be additional practical concerns associated with pedagogical implementation.

Addressing these concerns, Olson (2014), following the implementation of a VFP in a third semester university level Spanish course, assessed student perceptions of the VFP via a usability questionnaire. Overwhelmingly, students found the PRAAT software easy to use. Moreover, students felt that the VFP was a good method for conceptualizing pronunciation and would ultimately be beneficial to improving pronunciation. Olson concluded that although more streamlined, user-friendly visual feedback software should be an ongoing goal in the field, the current researcher-focused iterations are clearly usable in the L2 classroom.

### **Research Questions**

Considering the need for systematic methods of pronunciation instruction and the relevancy of the visual feedback paradigm for improving segmental production, the research questions of the current paper are twofold: (1) Does a visual feedback paradigm significantly improve L2 segmental production in a classroom setting? And (2) are L2 segmental production gains generalizable to similar contexts?

## **METHODOLOGY**

To answer the above questions, this paper presents a visual feedback paradigm tailored to the classroom setting, implemented with lower-level language learners. The benefits of the VFP on segmental pronunciation, limited to the production of intervocalic stop consonants, are addressed through acoustic analysis. Generalizability is addressed in two distinct ways: (a) Segmental generalizability, defined as gains made on non-target segments from training on target segments in a similar phonetic context (i.e., improvement for [β] production during training on [ð]); (b) Utterance-level generalizability, defined as gains made on segmental production in connected speech resulting from training on targets in isolation.

### **Participants**

Participants in the study were drawn from two third semester Spanish course sections at a large, Midwestern university. The third semester Spanish course is considered to be an intermediate-low level course, focused on the four basic language skills and on culture. As such, participating students have a basic knowledge of Spanish structures. One section ( $N = 26$ ) served as the experimental group, receiving

the visual feedback paradigm described below, while the other section ( $N = 24$ ) served as a control. The assignment of control or experimental group was determined at random.<sup>1</sup> The instructor, the same for both groups and semi-blind to the nature of the experiment, was informed that the visual feedback paradigm was being evaluated for inclusion in the curriculum. For both groups, the activities formed part of the curriculum and were thus obligatory. After the final activity, both groups were informed about the experiment and given the option to provide their data, via informed consent, for research purposes. All participants, from both the control and experimental groups chose to provide their data and participate in the study. Participants received no compensation for agreeing to provide their data.

All participants were native English speakers and reported growing up in English-dominant regions of the US, with little to no contact with Spanish in daily interactions. Three participants reported spending significant time, greater than one month, in a region where Spanish is a dominant or co-dominant language, and were subsequently eliminated from the analysis. Two participants reported having taken a phonetics course, thus were also eliminated from the analysis. Relevant for the current study, all participants placed into this course via a standardized placement test or as the result of successfully completing the previous course in the sequence, ensuring relatively similar proficiency levels. Those taking the placement test had some number of Spanish courses either during secondary education (i.e., high school) or at another university. Lastly, all participants reported their current use of Spanish in contexts outside the classroom, on a scale from 1 to 9 (1 = *Use Spanish daily*; 9 = *Never use Spanish*). Participants reported minimal use of Spanish outside the classroom ( $M = 7.3$ ,  $SD = 1.0$ ). In sum, all participants had English as their first language (L1), with no early experience speaking any language other than English, minimal current interaction with Spanish outside the classroom, and no formal phonetic training.

### Stimuli

Stimuli for the paradigm (pre-test, VFPs, and post-test) consisted of Spanish words containing the segments [β, ð, γ] in intervocalic position, produced in two contexts: in isolation and embedded in utterances. The segmental features studied here, commonly mispronounced by native English speakers learning Spanish, are the intervocalic voiced stops /b, d, g/. In English, the occlusive realizations of these phonemes [b, d, g] are produced in all phonological contexts. In Spanish, the voiced stop is produced as an approximant allophone [β, ð, γ] in intervocalic position (e.g., Hualde, 2005).<sup>2</sup> Native L1 English learners of Spanish, with the exception of advanced learners, almost exclusively produce the full stop [b, d, g] in intervocalic position (e.g., Lord, 2010), as a result of L1 transfer (e.g., Eckman, 2008).

The first set of stimuli consisted of a set of 15 words in isolation, used in each of the recordings (i.e., pre-test, 3 VFPs, delayed post-test) (Appendix A). There were five occurrences of each of the three segments [β, ð, γ]. All isolation tokens consisted of two or three syllable words. To control for the effect of stress on lenition (Ortega-Llebaria, 2004), each token was controlled for stress with the target phone occurring in an unstressed syllable, immediately preceded by a tonic syllable. In addition, each target was controlled with respect to the phonetic context, with an equal distribution of the five Spanish vowels (/i, e, a, o, u/) preceding the target segment. Owing to restrictive Spanish morphology, the vowel following the target phone was always a mid or low vowel. Table 1 shows sample target tokens.

Table 1. Sample Stimuli.

[β]		[ð]		[γ]	
<i>escribo</i>	I write	<i>comida</i>	food	<i>digo</i>	I say
<i>subo</i>	I go up	<i>ayudo</i>	I help	<i>jugo</i>	juice
<i>sabe</i>	he knows	<i>cansado</i>	tired	<i>paga</i>	he pays

Note: Spanish target tokens are indicated in italics. English translations are provided here for convenience.

Second, unique sets of nine novel words, also containing [β, ð, γ] in intervocalic position, embedded in utterances were also recorded during each session (Appendix B). The novel words were not part of the set of words in isolation that was included in the training. These sets of novel words in utterances were included to gauge the extent of carry over from the tokens addressed in isolation to tokens in connected speech (i.e., utterance-level generalizability). A different set of novel words was used in each session and none was included in any of the visual analyses. Parallel to the isolated word tokens, each embedded token was controlled for stress and phonetic context. Ordering of all stimuli was randomized.

A total of 3,656 tokens were coded for analysis (2,280 words in isolation and 1,376 novel words in utterances).<sup>3</sup>

### **Procedure**

The control group received typical pronunciation verbal feedback, predominantly using an ad-hoc approach of individual correction and correct pronunciation modeling, representative of the most common approaches found in the previous literature (e.g., Olson, 2014). The types of corrective feedback reported by the course instructor aligned with the feedback typology proposed by Lyster and Ranta (1997) and included explicit correction, recasts, clarification requests, repetition, elicitation, and metalinguistic feedback. This approach took place in an unstructured manner, and based on previous instructor self-reported accounts (Olson, 2014), it is estimated that such feedback accounted for less than a few minutes per class period. For both groups, no correction was given for the target segments and no explicit articulatory explanations were provided. In addition, the experimental and control groups received the same total minutes of classroom instruction (i.e., 150 minutes per week), with the control group being given a corresponding amount of time dedicated to socio-cultural aspects of the Spanish language relative to the experimental group's time spent on the VFP. Furthermore, while the text used for the course contains a limited number of references to pronunciation, principally via orthographic conventions, these were not included as part of the course requirements, assigned readings, curriculum, or assessments for either group. Participants in the control group were also asked to perform audio recordings of the same stimuli, at the same point in the course, for the pre-test and the three activities. They never received any feedback on their recordings, and did not participate in the VFP. With the exception of the VFP for the experimental group (detailed below), the remainder of the course design (i.e., instructor, curriculum, assessments, etc.) was identical for both groups.<sup>4</sup> Thus, while the experimental group participated in the visual feedback paradigm and the control group did not, the type and quantity of feedback, total instruction time, and all other aspects of the course were similar for the two groups.

Broadly, for the experimental group, the procedure consisted of a pre-test, three separate VFPs of approximately 20 minutes, and a delayed post-test conducted three and a half weeks after the final VFP. Assignments were assessed on a completion/non-completion basis. The instructor never gave feedback with regard to how well the students performed the assigned task or pronounced the target word. Crucial for the question of segmental generalizability, while participants recorded productions of all three intervocalic segments during each recording session, the visual feedback paradigm conducted in each session focused only on a single phone: VFP 1 focused on [ð]; VFP 2 focused on [γ]; VFP 3 focused on [β].<sup>5</sup> The pre-test, each VFP, and the delayed post-test each took place approximately three weeks apart.

Each of the three VFPs consisted of: (a) initial self-recording; (b) guided visual analysis; and (c) practice and re-recording. The visual feedback paradigm employed the speech analysis program PRAAT (Boersma & Weenink, 2011), which was selected based solely on availability (free and downloadable) and versatility (available for Macintosh, PC, and Linux). Generally, PRAAT allows for the recording of speech sounds and provides a visual representation of the acoustic signal through waveforms, spectrograms, and intonation contours. Prior to the pre-test, students were given a brief tutorial in the L2 on the mechanics of using PRAAT to record a sound file and produce the “visual picture” (i.e., waveform and spectrogram).

### *Pre-recording*

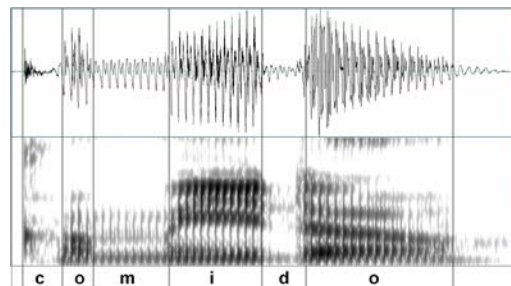
For each of the three VFPs, participants first recorded the given set of stimuli prior to the class meeting, generally on their personal computers. Participants were then instructed to bring to class the visual representation of the first four words produced in isolation from the list of stimuli. They were also directed to mark their images for the different “sounds/letters”, mainly through repeated listening to the recording and segments of the recording.

### *Self and Native Speaker Analysis*

In class, participants were instructed to: (a) visually examine their own productions, (b) visually analyze native speaker productions, (c) compare their own productions with those of a native Spanish speaker, and (d) compare three sets of spectrograms produced by native and non-native Spanish speakers.

Participants first answered a series of questions regarding their own productions, drawing particular attention to the segment in consideration and its production relative to preceding and following vowels. [Example \(1\)](#) provides sample guiding questions given to participants, originally given in the L2. Typical answers to this set of questions included that the “d” was much lighter in color than the vowels, roughly the same length, and was very different visually from the preceding and following vowel. These descriptions correspond to a full closure for the voiced occlusive and less energy across the frequency spectrum. [Figure 1](#) illustrates the spectrogram of the word *comido* (‘eaten’) produced by a native English speaker.

- (1) a. What are the visual characteristics of your “d”?
- b. Is your “d” darker or lighter than the sounds around it?
- c. Is your “d” easy or hard to distinguish from the sounds around it?



*Figure 1.* Sample spectrogram of the token *comido* (‘eaten’) produced by a native English speaker.

Participants then answered a parallel set of questions regarding a spectrogram and waveform produced by an age-matched, male native Spanish speaker (central peninsular dialect) ([Example 2](#)). Typical answers to this set of questions, provided by students in either the L1 or L2, included that the “d” produced by the Spanish speaker ([Figure 2](#)) was similar, or “slightly lighter” than the vowels, that it was generally shorter than the vowels, and that it was difficult to visually distinguish the “d” from the preceding and following vowels. These descriptions correspond to the approximant-like productions.

- (2) a. What are the visual characteristics of the Spanish speaker’s “d”?
- b. Is it darker or lighter than the sounds around it?
- c. Is it easy or hard to distinguish from the sounds around it?

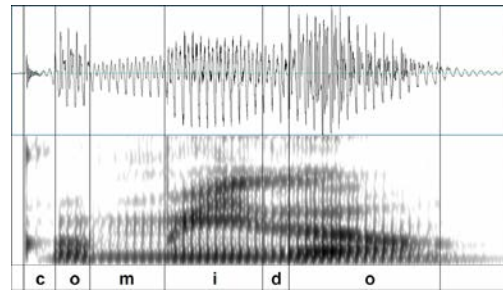


Figure 2. Sample spectrogram of the token *comido* ('eaten') produced by a native Spanish speaker.

Participants compared their own visual production with that of the native Spanish speaker and hypothesized about the auditory differences between the two productions (Example 3). Generally, participants noted their own stop productions were much easier to differentiate from the surrounding vowels than the native Spanish speakers and describe their own “d” as being more “closed”, “harder”, or less “connected” to the vowels than that produced by the native Spanish speaker. With respect to the auditory characteristics, participants hypothesized that the native Spanish speaker’s “d” was “noisier” and that there was “no silence”. These responses highlight the difference between the stop and approximant realizations of the target segment.

- (3) a. Describe the visual difference between your “d” and the “d” produced by a native Spanish speaker.
- b. What do you think the auditory difference is between your “d” and the “d” produced by the native speaker?

Finally, participants were given a series of pairs of spectrograms (Figure 3), each consisting of one produced by a native Spanish speaker and one produced by a native English speaker. Participants were asked to choose the one that was produced by the native English speaker.

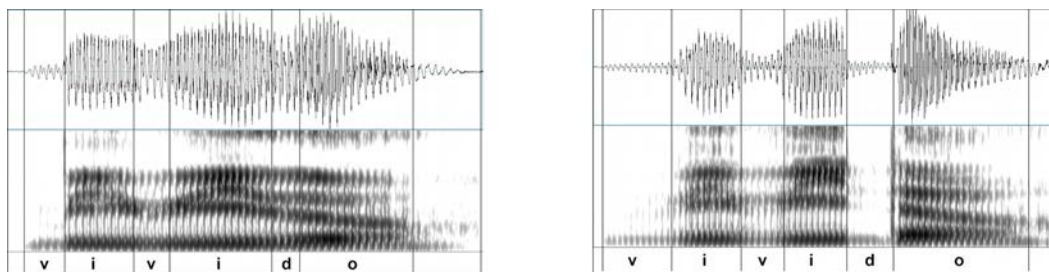


Figure 3. Pair of spectrograms: the left image corresponds to *vivido* ('lived') produced by a native Spanish speaker and the right image corresponds to *vivido* produced by a native English speaker.

### Practice and Re-recording

Following the VFP, participants were instructed to re-record the list of target words and stimuli, which were subsequently provided to the investigator for analysis. Participants were encouraged to record and visually examine their spectrograms before submitting their final recordings.

### Measurements

As an index of pronunciation accuracy for the intervocalic segment, and as an approximation of the degree of closure, each token was analyzed with respect to *intensity-ratio*. The intensity-ratio is defined as the minimum intensity (dB) of the target phone divided by the maximum intensity (dB) of the preceding



vowel (e.g., Ortega-Llebaria, 2004). To calculate the intensity-ratio, the investigator delimited the bounds of each token by hand, using PRAAT (Boersma & Weenink, 2011). Special attention was given to the second vowel formant in determining the bounds of the vowel, and by extension, the target segment. To standardize the measurement process, a script was used to automatically compute the intensity-ratio of each production. The intensity-ratio provides a measure of oral closure, with 1 representing a fully open, vowel-like consonant, and 0 representing a fully closed, silent stop.<sup>6</sup>

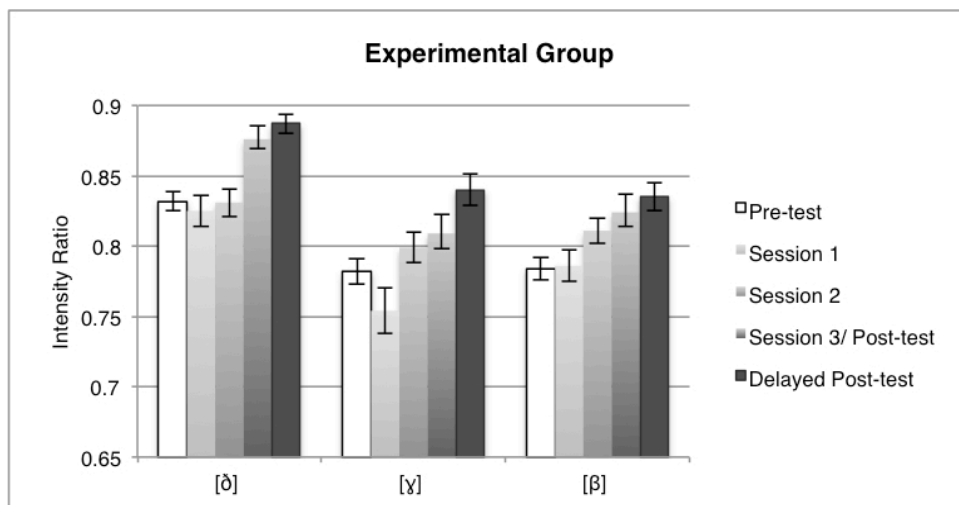
## RESULTS

### Words in Isolation

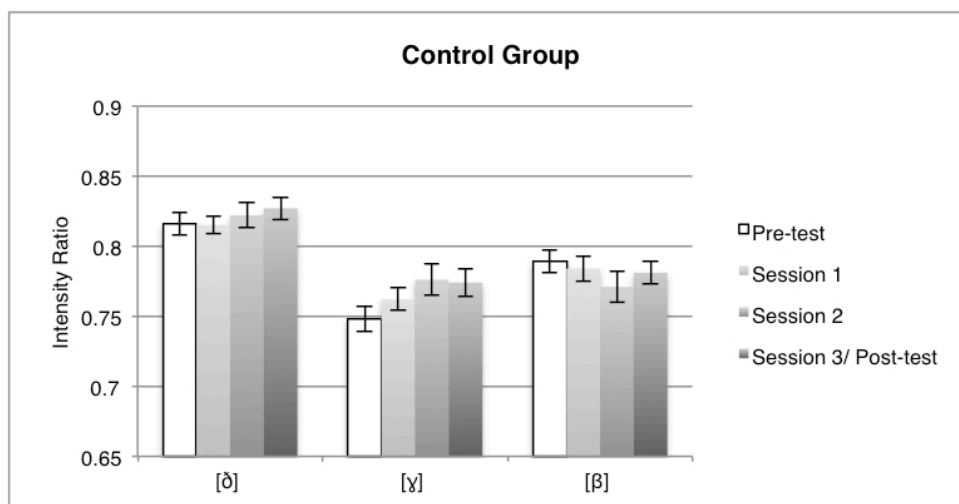
A univariate ANOVA was performed on the dependent variable of intensity-ratio, with main factors of *session* and *group*.<sup>7</sup> *Session* was defined as the specific recording session. For the experimental group, *session* refers to each VFP activity (i.e., pre-test, VFP 1, VFP 2, VFP 3, delayed post-test). For the control group, *session* refers to the corresponding recording of the target utterances. Given that Session 3 was the final recording activity, it is considered to be a post-test. *Group* consisted of the control or experimental group. Initial results revealed a significant effect of both *session* ( $F(4, 2271) = 18.72, p < .001, \eta_p^2 = .022$ ) and *group* ( $F(1, 2271) = 20.78, p < .001, \eta_p^2 = .010$ ), as well as a significant interaction between the two (*session*  $\times$  *group*,  $F(3, 2271) = 3.77, p = .010, \eta_p^2 = .005$ ).

A Tukey post-hoc analysis, with respect to the comparison between the control and experimental groups, demonstrates that there was no significant difference between the groups at the time of the pre-test (*diff.* = .014,  $p = .653$ ), demonstrating that the two groups were originally homogeneous with respect to the production of intervocalic stops in Spanish. Illustrating the impact of the training on the experimental group, the post-hoc comparison of the two groups following Session 3 (i.e., post-test) reveals a significant difference between the two groups (*diff.* = .042,  $p < .001, d = .485$ ). Analysis of the mean intensity-ratio shows that the productions of the experimental group are more native-like, with intensity-ratio values closer to 1 ( $M = .836, SD = .095$ ) than the control group ( $M = .794, SD = .098$ ). As such, the analysis reveals that the experimental group improved with respect to the control group.

A second layer of analysis focused on the improvement within each group, specifically a comparison of the pre-test and Session 3. The analysis (Tukey) showed no significant difference between the intensity-ratios of the control group at the pre-test and Session 3 (*diff.* = -.009,  $p = .942$ ). However, similar analysis for the experimental group returned distinct results, with significant differences found between the intensity-ratios at the pre-test ( $M = .799, SD = .071$ ) and Session 3 ( $M = .836, SD = .095$ ) (*diff.* = .037,  $p < .001, d = .446$ ). Also worth noting, is that for the experimental group, the effects of the training lasted into the delayed post-test, with no significant difference between Session 3 and the delayed post-test (*diff.* = .018,  $p = .790$ ).



*Figure 4a.* Mean intensity-ratio over time for the experimental group by segment. Error bars represent  $\pm 1$  SE.



*Figure 4b.* Mean intensity-ratio over time for the control group by segment. Error bars represent  $\pm 1$  SE.

Figures 4a and 4b illustrate the mean intensity-ratios for each group during the pre-test, Session 1, Session 2 and Session 3. The graph also includes the performance of the experimental group during the delayed post-test. Visual analysis shows a clear positive trend for the experimental group, with successively greater intensity-ratios produced over time, while the control group shows overall little variation. In sum, for words in isolation, while both groups performed similarly during the pre-test recording, the experimental group evidenced significant increases in their intensity-ratios for intervocalic stops, while the control group showed no differences in performance over the duration of the study.

### Generalizability

The second goal of this paper is to address the issue of the generalizability of the training paradigm. Specifically, generalizability was assessed both at the segmental level (i.e., improvement for non-target phones during VFP) and the utterance level (i.e., improvements on productions in connected speech resulting from training on isolated words). Each issue is addressed here in turn.

### Segmental Generalizability

To address the question of whether improvement occurs only during a VFP focused on a particular segment or if such training may extend to other segments in the same phonetic environment, a post-hoc analysis was used to investigate the effect of visual feedback on the target phone. Specifically, the analysis focused on the difference in performance for each phone ([β], [ð], [ɣ]) in the recording immediately prior to and immediately following the visual feedback activity focused on that segment. For example, as the alveolar [ð] was the subject of training during VFP 1, analysis focused on a comparison of the intensity-ratio of [ð] in the pre-test and VFP 1. Analysis, using a non-paired t-test with unequal variance on collapsed data for all three phones immediately prior to and following the VFP focused on that phone (VFP 1- [ð]; VFP 2- [ɣ]; VFP 3- [β]) revealed no significant effect of training ( $t(422) = -1.337$ ,  $p = .182$ ).

Subsequent analysis focused on differences in performance either before the visual feedback paradigm focused on a given segment or after. Exemplifying the analysis of a phone prior to the visual feedback paradigm, as the labial [β] was the focus of the training in VFP 3, analysis centers on changes in intensity-ratio for [β] between the pre-test and VFP 2. Analysis of the effects after training would be correspondingly exemplified by the difference in performance on the alveolar [ð], the focus of the training in VFP 1, following VFP 1 and VFP 3. While analysis for segments prior to training revealed no effect ( $t(239) = .363$ ,  $p = .716$ ), there was a significant effect on intensity ratio after training ( $t(235) = -2.617$ ,  $p = .009$ ,  $d = .224$ ).

Figure 5 illustrates the pre- and post- collapsed mean intensity-ratio for all three phones: (a) prior to the visual feedback activity focused on the given segment; (b) during the visual feedback activity; and (c) after the visual feedback activity for a given segment. Visual inspection confirms the above findings. While production did not improve until the VFP focused on a given segment, there was a trend towards improvement during the activity, and significant improvement during subsequent VFPs focused on other segments. Given the overall improvements found for the experimental group, we can conclude that the marginal gains made during, coupled with greater gains after the visual feedback activity, combined to create statistically significant gains.

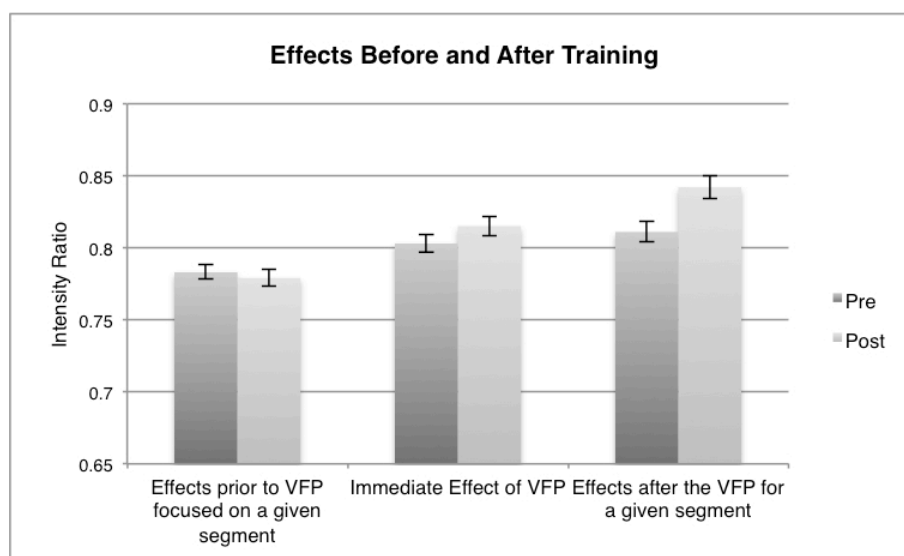


Figure 5. Mean intensity ratio, collapsed for all segments: (a) Prior to the VFP for a given phoneme (i.e., during VFPs focused on other target segments); (b) During the VFP; (c) Following the VFP for a given segment. Error bars represent +/- 1 SE.

### Utterance-level Generalizability

The final analysis focused on the effect of training on student productions in a ‘more difficult’ context. Specifically, while training focused solely on words in isolation, the question remained if such gains would carry over to novel words, not included in the words-in-isolation training, embedded in utterances. Paralleling the analysis performed for the words in isolation, a similar ANOVA was conducted with main factors of *session* and *group*, revealing a significant effect of both *session* ( $F(4, 1366) = 7.13, p < .001, \eta_p^2 = .018$ ) and *group* ( $F(1, 1366) = 8.11, p = .004, \eta_p^2 = .009$ ), but no interaction between the two ( $session \times group, F(3, 1366) = .97, p = .404$ ). Overall, as illustrated in Figure 6, there was a general increase in the intensity-ratios over time for the experimental group. Worth noting, post hoc testing (Tukey) revealed that while this effect was not statistically significant following Session 3 (pre-test vs. Session 3,  $p = .417$ ), it was significant at the time of the delayed post-test (pre-test vs. post-test,  $diff. = .043, p = .045, d = .400$ ).

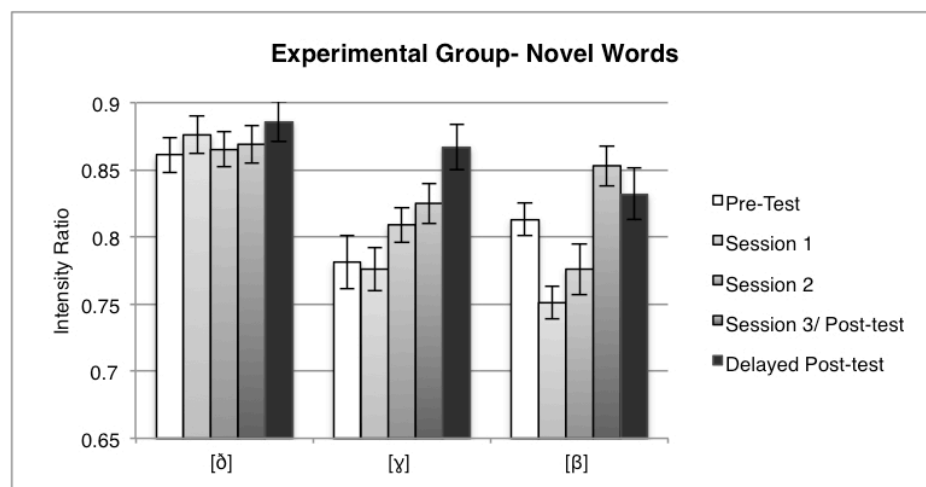


Figure 6. Mean intensity ratio over time for novel tokens embedded in utterances. Error bars represent +/- 1 SE.

## DISCUSSION

The current study answers previous calls in the literature for both controlled investigations of the effects of a visual feedback paradigm at the segmental level and empirical evidence for the pedagogical effectiveness of VFP in the L2 classroom. Results demonstrated the potential effectiveness of a visual feedback paradigm in the L2 classroom for the teaching of a segmental feature. Comparisons between the experimental and control groups, as well as the pre- and post-tests of the experimental group, demonstrated that the visual feedback paradigm results in significantly improved, more native-like, productions of the intervocalic stops for L1 English learners of Spanish. Furthermore, this study suggests that while a single 20-minute iteration of a VFP may create marginal gains, significant gains occur as the result of incremental gains both during and after the VFP focusing on a given target. Lastly, results with respect to the generalizability to connected speech are promising, but warrant further investigation.

As a first contribution, the current paradigm shows that VFP can be successful at the segmental level, answering previous calls for visual feedback beyond the prosodic level (Chun, 2007). At the segmental level, previous research has focused predominately on vowel production (Carey, 2004; Okuno, 2013; Ruellot, 2011; Saito, 2007; for geminates: Motohashi-Saigo & Hardison, 2009), potentially owing to the effect of vowel production on accentedness and intelligibility. The current study represents one of the first investigations of the effect of visual feedback on consonant production. While previous results concerning

the effectiveness of a VFP on vowel production have yielded mixed results, the current results for intervocalic stops suggest the potential of this technology for segmental instruction.

As a whole, research on the effectiveness of the VFP has produced generally positive results for suprasegmental instruction. At the segmental level, while results have been less consistent, the current study showed significant effects for intervocalic stop production. Attempting to account for these differences (i.e., positive vs. inconsistent results), it is worth mentioning the potential difference in the intuitive nature of visual interpretation of suprasegmentals, vowels, and consonants. Specifically, intonation contours may have the benefit of being intuitively understandable to novice learners (Léon & Martin, 1972), with variations in the contour relating directly to variations in pitch. In contrast, it is not clear that vowel features, like correlation between formants and tongue position, can be considered as intuitive. Without some additional training, perhaps more suited for a pronunciation or phonetics course (Lord, 2010; Ruellot, 2011), the use of spectrograms for visual feedback may be constrained by what can be intuitively understood by learners. Consonants, varying widely in the relevant features visible on a spectrogram, may represent a middle ground. While some features, like intensity and duration, may be easily perceived and understood, others such as nasality or rhoticity, may be less intuitive. For researchers and instructors, it is worth considering which features are most likely to be successful within a visual feedback framework.

Addressing the pedagogical need, the results of the current study suggest that the VFP, historically the focus of laboratory-based training or incorporated within a larger phonetic course, is potentially suitable for the general skills L2 classroom. The VFP implemented here, based largely on that suggested by Olson (2014), required no special access to technology during the class itself, and relegated much of the pronunciation practice as homework. That is, not only can such a VFP produce quantifiable benefits for learners, it allows for practical integration into the L2 general skills classroom without any special modifications or large amounts of in-class time.

Lastly, it is worth considering not only *how* the VFP presented here and elsewhere is successful, but also potentially *why* it may be successful. As noted by Derwing and Munro (2005), just as in the domains of syntax and morphology, learners often need help “noticing what they are doing” (p. 387). As suggested by Schmidt (1990), noticing is an important step in processing the relevant input, applicable at all levels of language learning (c.f. acquisition), including the lexicon, syntax, and phonology (see also Gattegno, 1987). This point echoes the findings of the effect of perceptual training on phonetic production (e.g., Bradlow et al., 1997), with the ability to distinguish two sounds having a positive effect on production accuracy. Within the classroom, listening and repeating is one of the most common methods for teaching pronunciation, but learners may not notice the differences with this method, unless explicitly directed to do so by the instructor (Miller, 2012). Moreover, auditory perception is limited by the phonotactic constraints of the L1 (e.g., Flege & Wang, 1989), such that phonetic distinctions not relevant to the L1 may be collapsed into a single perceptual category, impeding perceptual distinction in the L2. As such, the effectiveness of the visual feedback paradigm may lie in its ability to allow learners to notice the differences between their own productions and those of a native speaker. Visual perception may provide a second modality to facilitate noticing, particularly relevant for cases in which auditory perception is limited. Thus, the benefits of the VFP evident through empirical research, may be explained within the framework of noticing and discrimination.

## **CONCLUSIONS AND FUTURE RESEARCH**

The current study has attempted to address both the lack of systematic pedagogical tools for pronunciation instruction and the gap between laboratory and classroom research by presenting a visual feedback paradigm for pronunciation instruction in the lower-level general skills classroom. The results demonstrated that a VFP is more effective at segmental pronunciation instruction than the traditional ad-hoc approach. In addition, the findings suggest that a VFP results in incremental gains both at the time of

the intervention, as well as following the intervention. Furthermore, gains may be potentially generalizable from isolated word training to connected speech production.

Moving forward, subsequent research should compare the effectiveness of a VFP with various methodologies (e.g., ad-hoc approach, explicit articulatory instruction, metalinguistic explanation, etc.) in order to maximize pronunciation gains. In addition, although the commonly mispronounced segment under focus in the current study provides a good test case for the VFP, it should be acknowledged that while it may impact the degree of accentedness, it is not clear that this subphonemic phenomenon would impact either comprehensibility or intelligibility (Derwing, Munro, & Wiebe, 1998). In light of this limitation, focus should be turned to segments and suprasegmentals shown to most impact intelligibility (Chun, 2007; Munro & Derwing, 1995). Finally, the current study used only one of many possible pedagogical implementations of the VFP. Future creativity on the part of instructors and investigation by researchers will serve to improve the methods presented here. While the visual feedback paradigm shows great promise in the L2 classroom, much work remains to be done to implement this technology in a systematic way that maximally benefits students.

---

### APPENDIX A: Stimuli: Words in Isolation

[β]		[ð]		[ʎ]	
<i>escribo</i>	I write	<i>comida</i>	food	<i>digo</i>	I say
<i>debo</i>	I should	<i>queda</i>	he remains	<i>llega</i>	he arrives
<i>sabe</i>	he knows	<i>cansado</i>	tired	<i>paga</i>	he pays
<i>escoba</i>	broom	<i>todo</i>	everything	<i>ahogo</i>	he drowns
<i>subo</i>	I go up	<i>ayudo</i>	I help	<i>jugo</i>	juice

Note: Spanish target tokens are indicated in italics. English translations are provided here for convenience

### APPENDIX B: Stimuli: Novel Words in Utterances

#### Visual Feedback Paradigm 1

Target Segment	Utterance
[β]	Mi habitación está <b>arriba</b> de la cocina. Siempre le <b>daba</b> dinero a mi sobrino para navidad. La suegra siempre tira papel en el <b>cu</b> bo de basura.
[ð]	María va con su <b>marido</b> al Festival de la Calle Ocho. Tú tienes <b>dua</b> s sobre nuestras vacaciones en Cozumel. El hermano de mi esposo es mi <b>cuña</b> do favorito.
[ʎ]	Cuando se porta mal, la madre castiga al niño. Una ensalada tiene <b>le</b> chuga y tomate. Mi padre va al <b>la</b> go para pescar.

Note: Target sounds indicated in bold.

**Visual Feedback Paradigm 2**

Target Segment	Utterance
[β]	Cuando mi hermano hace una torre con bloques, yo derribo la torre. Cuando buceas, es importante que subas al superficie en una hora. Mis compañeros tiraban la pelota mucho en su tiempo libre.
[ð]	Mi madre ha cosido muchos guantes. Mi padre es muy peludo con mucho pelo en la espalda. Mi hermana siempre es muy callada y tonta.
[ʎ]	Mi compañero liga con muchas chicas porque es estrella de fútbol. Me gusta comer pechuga de pollo en un restaurante cerca del Parque Chapultepec. Vamos a tomar un trago un el bar y ver la lucha libre en la televisión.

Note: Target sounds indicated in bold.

**Visual Feedback Paradigm 3**

Target Segment	Utterance
[β]	No creo que recibas una A en la clase de arquitectura. Quiero que subas al tercer piso donde tenemos otro baño. No sé usar la máquina porque mi madre siempre lavaba mi ropa.
[ð]	Una vez en mi vida quiero visitar la Casa Batilló de Gaudí. No quiero que estés tan testarudo conmigo. Siempre he cortado el césped los domingos.
[ʎ]	Deja de tomar cerveza, no quiero que tengas una barriga grande. La piscina tiene una fuga y por eso llamé a la plomera. Por favor, apaga la chimenea.

Note: Target sounds indicated in bold.

**Post-test**

Target Segment	Utterance
[β]	No estoy seguro que mi novio escriba un poema para mi. Cuando vamos a la playa, espero que no haya nubes en el cielo. Mi madre siempre cocinaba pollo con ajo.
[ð]	Me gusta comer batidos con fresa y plátano. Celebrando el Día de Garífuna en Honduras, bailo y sudo mucho. Me gusta comer la carne asada con salsa en pupusas.
[ʎ]	Es importante que diga la verdad. En pocos países se come sopa de tortuga como un plato tradicional. No creo que el cocinero haga flan.

Note: Target sounds indicated in bold.

## NOTES

1. As with most classroom research, assignment of participants to the Experimental and control groups is only quasi-random. To further validate the homogeneity of the two groups, comparisons of pre-test performance were conducted (see Results), demonstrating no difference between the groups with respect to the target features.
2. The approximant allophones are represented here as [β, ð, ɣ] following conventions used in both Hualde (2005) and Hammond (2001). As a point of comparison with English, Hualde (2005) notes that the Spanish variants can be realized with very little constriction and no frication, as opposed to the English [ð], which is a true fricative. As noted by one reviewer, the corresponding status of the English [ð] may explain the superior performance on this segment in Spanish relative to [β, ɣ], which have no English counterparts.
3. The total number of tokens analyzed was impacted by a number of limitations of classroom research including participants missing treatments and poor (at-home) recording quality. In total, 21.4% of the data was missing (control: 17.9%; experimental: 25.5%).
4. While the design of the classes themselves may have been identical, it should be noted that, owing to external factors (e.g., student personalities and motivation), no two classes are ever identical.
5. While the order of phones addressed was determined randomly, given classroom constraints, it was not possible to counterbalance the ordering of the VFP.
6. Given that none of the recordings took place in a sound-dampened environment, intensity measurements may have been impacted. However, any background noise would effectively raise the intensity for the consonant, creating an artificially elevated, more native-like, intensity-ratio. As such, the current study may actually *underestimate* the effect of the VFP on segmental production.
7. While the initial experimental design lends itself to a repeated measures ANOVA, missing recordings for several students, resulted in an unbalanced data set that precluded the use of a repeated measures ANOVA. While a repeated measures ANOVA would account for the between-subjects variation, by including such variation a standard ANOVA may underestimate significance.

---

## ACKNOWLEDGEMENTS

I would like to thank Heather Offerman for her dedicated efforts on this project, three anonymous reviewers for their comments, as well as all of the participants for their collaboration. All errors, of course, remain my own.

---

## ABOUT THE AUTHOR

Daniel J. Olson is an Assistant Professor of Spanish and Linguistics at Purdue University. His research interests include experimental approaches to bilingual phonetics, in both early and late bilinguals, the storage and acquisition of distinct phonetic inventories, and code-switching.

**E-mail:** [danielolson@purdue.edu](mailto:danielolson@purdue.edu)

---

## REFERENCES

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, 20, 51–62.



- Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, 11(4), 5–22.
- Archibald, J. (1998). *Second language phonology*. Amsterdam, The Netherlands: John Benjamins.
- Arteaga, D. (2000). Articulatory phonetics in the first-year classroom. *The Modern Language Journal*, 84, 339–354.
- Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (version 5.2.16) [computer software]. Available from [www.praat.org](http://www.praat.org).
- Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985.
- Breitkreutz, J., Derwing, T., & Rossiter, M. (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.
- Burns, A. (2006). Integrating research and professional development on pronunciation teaching in a national adult ESL program. *TESL Reporter*, 39, 34–41.
- Castino, J. (1996). Impact of a phonetics course on FL learners' acquisition of Spanish phonology. *Selecta: Journal of the Pacific Northwest Council on Foreign Languages*, 17, 55–58.
- Carey, M. (2004). CALL Visual feedback for pronunciation of vowels: Kay Sona-Match. *CALICO Journal*, 21(3), 1–32.
- Chun, D. (1989). Teaching tone with microcomputers. *CALICO Journal*, 7(1), 21–47.
- Chun, D. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology*, 2(1), 61–77. Retrieved from <http://llt.msu.edu/vol2num1/article4/index.html>
- Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam, The Netherlands: John Benjamins.
- Chun, D. (2007). Come ride the wave: But where is it taking us? *CALICO Journal*, 24(2), 239–252.
- Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, 18, 53–70.
- Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Ames, IA: Iowa State University.
- 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?* Edmonton, AB: Prairie Metropolis Centre.
- Derwing, T. (2010). Utopian goals for pronunciation teaching. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 24–37). Ames, IA: Iowa State University.
- Derwing, T., Diepenbroek, L. & Foote, J. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, 30(1), 22–44.
- Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching: A research-based

approach. *TESOL Quarterly*, 39, 379–397.

Derwing, T., Munro, M., & Wiebe, G. (1997). Pronunciation instruction for fossilized learners: Can it help? *Applied Language Learning*, 8, 185–203.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410.

Derwing, T., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13, 1–17.

de Bot, K. (1980). Evaluation of intonation acquisition: A comparison of methods. *International Journal of Psycholinguistics*, 7, 81–92.

de Bot, K. (1983). Visual feedback of intonation: Effectiveness and induced practice behavior. *Language and Speech*, 26, 331–350.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16, 71–77.

Eckman, F. (2008). Typological markedness and second language phonology. In J. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 95–116). Philadelphia, PA: John Benjamins.

Elliott, R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal*, 79, 530–542.

Elliot, R. (1999). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, 80(1), 95–108.

Flege, J. & Wang, C. (1989). Native language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, 17, 299–315.

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

Foote, J., Trofimovich, P., Collins, L., & Soler Urzúa, F. (2013). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, 1–16.

Gattegno, C. (1987). *The science of education. Part I: Theoretical considerations*. New York, NY: Educational Solutions.

González-Bueno, M. (1997). The effects of formal instruction on the acquisition of Spanish stop consonants. In W. Glass & A. T. Perez-Leroux (Eds.), *Contemporary perspectives on the acquisition of Spanish, Vol. 2: Production, processing, and comprehension* (pp. 57–75). Somerville, MA: Cascadilla.

Hammond, R. (2001). *The sounds of Spanish: Analysis and application*. Somerville, MA: Cascadilla.

Hardison, D. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, 8(1), 34–52. Retrieved from <http://llt.msu.edu/vol8num1/hardison/default.html>

Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentation with transient visual feedback. *Language Learning and Technology*, 13(3), 32–50. Retrieved from <http://llt.msu.edu/vol13num3/hincksedlund.pdf>

Hualde, J. I. (2005). *The sounds of Spanish*. New York, NY: Cambridge University Press.

Kay Elemetrics. (1986). *Visi-Pitch/Apple applications manual*. Pine Brook, NJ: Kay Elemetrics.

- Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12(2), 137–156.
- Leather, J. (2000). Second-language speech research: An introduction. *Language Learning*, 49, Supplement 1, 1–56.
- Léon, P., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *The Modern Language Journal*, 56(3), 139–144.
- Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33, 37–63.
- Levis, J., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, 12(2), 13–20.
- Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32, 505–524.
- Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 88(3), 557–567.
- Lord, G. (2010). The combined effects of instruction and immersion on second language pronunciation. *Foreign Language Annals*, 43(4), 488–503.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 16(1), 37–66.
- MacDonald, S. (2002). Pronunciation: Views and practices of reluctant teachers. *Prospect*, 17(3), 3–18.
- Major, R. (1998). Interlanguage phonetics and phonology: An introduction. *Studies in Second Language Acquisition*, 20(1), 131–37.
- Miller, J. S. (2012). Teaching pronunciation with phonetics in a beginner French course: Impact of sound perception. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 109–123). Ames, IA: Iowa State University.
- Molholt, G. (1988). Computer assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, 22(1), 91–111.
- Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates training with waveform displays. *Language Learning & Technology*, 13(2), 29–47. Retrieved from <http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf>
- Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97.
- Murphy, D. (2011). An investigation of English pronunciation teaching in Ireland. *English Today*, 27(4), 10–18.
- Okuno, T. (2013). *Acquisition of L2 vowel duration in Japanese by native English speakers*. (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.
- Olson, D. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in L2 pronunciation instruction. *Hispania*, 97(1), 1–25.
- Ortega-Llebaria, M. (2004). Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/. In T. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 237–253). Berlin, Germany: Mouton de Gruyter.
- Pennington, M., & Ellis, N. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, 84(3), 372–389.

- Ruellot, V. (2011). Computer-assisted pronunciation learning of French /u/ and /y/ at the intermediate level. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 199–213). Ames, IA: Iowa State University.
- Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation teaching in EFL settings: The case of English vowels and Japanese learners of English. *The Linguistics Journal*, 3(3), 16–40.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–159.
- Seferoğlu, G. (2005). Improving students' pronunciation through accent reduction software. *British Journal of Educational Technology*, 36(2), 303–316.
- Setter, J., & Jenkins, J. (2005). State of the art review: Pronunciation. *Language Teaching*, 38, 1–17.
- Simões, A. (1996). Assessing the contribution of instructional technology in the teaching of pronunciation. *Proceedings of the International Conference on Spoken Language Processing* (pp. 1461–1464). Philadelphia, PA.
- Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26, 233–253.
- Thomson, R. (2012). Improving L2 listeners' perception of English vowels: A computer mediated approach. *Language Learning*, 64(4), 1231–1258.
- Wang, X., & Munro, M. (2004). Computer-based training for learning English vowel contrasts. *System*, 32, 539–552.
- Weltens, B., & de Bot, K. (1984). The visualization of pitch contours: Some aspects of its effectiveness in teaching foreign intonation. *Speech Communication*, 3(2), 157–163.
- Zampini, M. (1996). Voiced stop spirantization in the ESL speech of native speakers of Spanish. *Applied Psycholinguistics*, 17, 335–354.