Purdue University
Purdue e-Pubs

School of Languages and Cultures Faculty Publications

School of Languages and Cultures

7-12-2016

Visual Feedback and Second Language Segmental Production: the Generalizability of Pronunciation Gains

Daniel J. Olson
Purdue University, danielolson@purdue.edu

Heather Michelle Offerman
Purdue University, hofferma@purdue.edu

Follow this and additional works at: https://docs.lib.purdue.edu/lcpubs

Visual feedback and second language segmental production: The generalizability of pronunciation gains

Heather Offerman and Daniel J. Olson

Purdue University
640 Oval Dr., West Lafayette, IN 47907, USA
hofferma@purdue.edu, danielolson@purdue.edu

Abstract
While a number of researchers have noted the lack of research on pronunciation instruction, relative to other aspects of language (i.e. syntax), pronunciation has been shown to be crucial for facilitating intelligible and comprehensible second language (L2) productions. Addressing the need for empirically tested pedagogical methods, the current study considers the use of a classroom-based visual feedback paradigm for the instruction of a segmental feature, namely voice onset time, which has been shown to be a distinctive marker of accent for English-dominant L2 learners of Spanish. In addition, this study examines the potential generalizability of gains made through the visual feedback paradigm, assessing whether gains made in controlled reading tasks (i.e. carrier sentences) will extend to more continuous and spontaneous speech. The results demonstrate significant improvements in voice onset time produced by participants following the visual feedback paradigm, relative to a control group. Furthermore, while the visual feedback training was limited to short, controlled utterances (i.e. carrier sentences), benefits were observed for more continuous and spontaneous speech.

1. Introduction

Within the field of second language acquisition, pronunciation has received relatively little attention, particularly with respect to the more prominent areas of morphology and syntax (Deng et al., 2009). However, highlighting the importance of pronunciation in language acquisition, research has shown that pronunciation has clear impacts on conveying meaning effectively (i.e. comprehensibility) and efficiently (i.e. intelligibility), and that accentedness in a second language (L2) may drive unwanted negative evaluations of L2 speakers by native speakers of the target language. As such, a number of researchers have begun to call for empirically-based research on pronunciation instruction (e.g. Wang & Munro, 2004). This lack of research can also be seen in the general lack of systematic pedagogical materials for pronunciation training, effectively reinforcing the "marginalized" nature of pronunciation in the second language classroom (Derwing & Munro, 2005, p. 382). In spite of this general lack of attention on pronunciation, students desire to learn more about pronunciation and believe that it is a critical component of L2 learning (Levis & Grant, 2003).

A growing body of research has begun to investigate new methods of instruction for L2 pronunciation, one method of particular interest has come to the forefront, combining pronunciation instruction with speech analysis technology. Specifically, visual feedback paradigms, seeking to aid learners in not only hearing their errors, but visualizing these errors, have been the subject of interest and investigations for several decades (for review see Chun, 1989). More recently the effectiveness of visual feedback for training at the segmental level (i.e. vowels and consonants) has been considered, albeit with somewhat varying results (e.g. Saito, 2007; Ruellot, 2011). While there is tacit support for the utility of visual feedback at the segmental level, further research on a variety of features is needed to confirm these findings. Furthermore, the support for the effectiveness of visual feedback has seemingly ignored the issue of generalizability, or whether gains made during training in restricted contexts may extend to more naturalistic speech.

Addressing these gaps, the current study investigated the effectiveness of visual feedback in aiding English-speaking L2 learners of Spanish to produce more target-like pronunciation of a segmental feature not previously addressed in the visual feedback literature, namely the voiceless stops (i.e. /p, t, k/). The visual feedback paradigm was used in order to highlight differences in voice onset time (VOT) values in English as opposed to Spanish, through comparisons of spectrograms and sound waves. Furthermore, the present study also sought to determine whether or not L2 learners were able to not only produce the target-like productions of /p, t, k/ in carrier utterances, but also in continuous controlled speech and spontaneous speech. Broadly, the positive results have both pedagogical and theoretical implications, highlighting the potential utility of visual feedback for pronunciation training.

2. Review of the Literature

2.1. Pronunciation: Intelligibility, Comprehensibility, & Accentedness

As L2 learners develop pronunciation skills, they are faced with issues of intelligibility, comprehensibility, and accented speech. *Intelligibility* is broadly defined as the extent to which an utterance is actually understood by the listener (Derwing & Munro, 2005). For L2 learners, many researchers have shown a link between pronunciation and intelligibility, with non-target-like pronunciation leading to decreased intelligibility (Derwing & Munro, 1997; Derwing, Munro, & Wiebe, 1998; Munro & Derwing, 1995, 1999; Derwing & Munro, 2009; Levis 2005; Levis & Grant, 2003; Derwing & Munro, 2009; Simões, 1996; Sturm, 2013). While intelligibility is an indicator of how much is understood, *comprehensibility* can be seen as degree of ease or difficulty with which an utterance can be understood (Derwing, Munro & Wiebe, 1998). In the case of non-native speakers (NNS), learning or acquiring the ability to produce intelligible and comprehensible communication has been discussed as a primary goal.

As defined by Derwing & Munro (2009), *accentedness* can be described as "how different a pattern of speech sounds to a local variety" (p. 478). Although accentedness does

not always correlate directly with intelligibility or comprehensibility for a NS listener (Derwing & Munro, 1997), it can impact NS perceptions about the NNS (Derwing & Munro, 2009; Gluszek & Dovidio, 2010; Kim, Wang, Deng, Alvarez, & Li, 2011; Purkiss, Perrewé, Gillespie, Mayes, & Ferris, 2006). Munro, Derwing, and Sato (2006); for example, a non-native-like accent may cause a NS to determine that the NNS is ignorant in their L2.[1] Furthermore, if a foreign accent is detected, NSs may determine that a NNS is not fluent in the target language, despite the NNS's use of correct syntax or grammar (Gluszek & Dovidio, 2010).

In short, for L2 speakers, pronunciation represents a major component for both expressing an easily understandable message (i.e. intelligibility and comprehensibility), as well as shaping listeners' perceptions (i.e. accentedness). Given the potential impacts on intelligibility, comprehensibility, and accentedness, pronunciation instruction is essential in aiding L2 learners to achieve communicative goals.

2.2. Approaches to L2 Pronunciation Instruction

Although second language pronunciation clearly has ramifications for intelligibility, comprehensibility, and accentedness, relatively little attention has been given to the teaching of pronunciation (Arteaga, 2000; Derwing & Munro, 2005; Elliott, 1995, 1997; Isaacs, 2009; Lord, 2005; Saalfeld, 2011; Saito, 2011, 2013; Simões, 1996), particularly in comparison with grammatical and syntactic features. Pennington and Richards (1986), for example, explain how pronunciation has been set aside as mere "linguistic competence" rather than "communicative competence" in more recent methods of instruction (p. 207). In other words, learning about pronunciation has been considered to be too detailed and too advanced for L2 learners, and

---

[1] Native listeners have also been shown to judge different and non-standard varieties (e.g. social, geographical, racial, etc) of their own language (for Spanish see Blas Arroyo, 1999), attributing both positive and negative characteristics to different varieties. While voice onset time, the focus of the current study, has been shown to vary minimally across various varieties of Spanish (e.g., Williams, 1977), it is worth considering and discussing with students the implications of the variety of any language used in pronunciation instruction.

they are taught mainly to focus on communicating an idea by using as much implicitly acquired knowledge as possible.

Beginning to address this need, there is a growing body of research that has shown a clear benefit for incorporating pronunciation instruction in the L2 classroom. Much of this work has addressed the potential impact of a variety of pedagogical or laboratory-based interventions, including auditory exposure (Neufeld, 1978), auditory discrimination training (Rosenman, 1987), awareness training (Pennington & Ellis, 200) explicit articulatory instruction (Castino, 1996; González-Bueno, 1997), and a variety of mixed methods (Couper, 2003, 2006; Derwing & Rossiter, 2003; Elliot, 1997; Gozález-Bueno, 1997; Lord, 2005; Santos Maldonado, 1994; for explicit instruction and immersion see Lord, 2010). For example, explicit articulatory instruction, in which students focus on learning parts of the vocal apparatus, has been shown to significantly improve pronunciation of certain phonetic features (Derwing, Munro, & Wiebe, 1997; Saito, 2012). Other researchers have taken a more multi-faceted approach, coupling explicit articulatory production with other pedagogical methods such as native speaker modeling (Saito, 2013), or general discussion of phonemic systems and allophonic distributions (Elliot, 1997; Lord 2005; Sturm, 2013). While a few studies have found no effects of pronunciation training (e.g. Macdonald, Yule, & Powers, 1994; Suter, 1976), as a whole, this body of research has shown significant gains resulting from pronunciation instruction.

This general lack of focus on pronunciation instruction can be seen both in research and pedagogical implementations. A number of studies have shown that while language instructors desire to include pronunciation in their classrooms (Levis & Grant, 2003), many are not well equipped with the skills to implement a structured method (Breitkreutz, Derwing, & Rossiter, 2002; Burgess & Spencer, 2002; Burns, 2006; Derwing & Munro, 2005). Specifically, many instructors have reported relying on their own intuitions when explaining pronunciation (Foote, Holtby, & Derwing, 2011; Levis, 2005), and generally approaching pronunciation in an ad-hoc

rather than systematic fashion (Olson, 2014a). Furthermore, recent examinations of L2 textbooks, which are often the focal point for teaching practices and in-class activities (Thomson & Derwing, 2004), have found a lack of structured pronunciation activities (Derwing, Diepenbroek, & Foote, 2012; Levis, 1999; Foote, Trofimovich, Collins, & Soler Urzúa, 2013; Rossiter, Derwing, Manimtim, & Thomson, 2010) and inaccurate, incomplete, or minimal information (Arteaga, 2000; Miller, 2012). In summary, there is a clear need for concrete, research-based, and goal-oriented pronunciation practices and activities in the second language classroom.

2.3. Technology and Phonetics Instruction

Within the relatively limited research on methods for pronunciation instruction, the use of technology, and specifically visualization, has begun to garner considerable support. Broadly, speech analysis technology that provides a visual representation of speech features (e.g. intonation, intensity, formant transitions, etc.) has been considered for its potential as a methodology for pronunciation instruction. Pedagogical or training implementation of visual representation has taken a number of different forms, including providing a visual model of NS pronunciation (e.g. de Bot, 1980), providing audio-visual modeling of NS productions (e.g. de Bot, 1983), and (nearly) simultaneous representations of L2 participant-produced speech compared with NS models (e.g. Olson, 2014a). In each case, the broad goal is for L2 learners to be presented with visual representations of NS productions, compare their own productions to those of the NS, and through noticing differences, improve pronunciation of the target feature.

Many of the first studies that addressed the efficacy of visual feedback focused primarily on suprasegmental features, namely intonation contours (Anderson-Hsieh, 1992, 1994; Chun, 1989, 1998, 2002; Molholt, 1988). In a seminal series of studies, de Bot and colleagues investigated the role of visual modeling and visual feedback on the production of

intonation patterns by L2 speakers (deBot & Mailfert, 1982; deBot, 1983; Hardison, 2004; Weltens & deBot, 1984). Using early software that created a line-drawing of the rise and fall of intonation (e.g. *Visipitch*: Kay Elemetrics, 1986), L2 learners were presented with their own intonation contours and as well as intonation contours of a NS of the target language. After comparing their own productions with those of a native speaker, this method also allowed L2 learners the opportunity to attempt to match the NS productions. As such, this visualization provided learners the opportunity to judge visually the degree of accuracy of their productions, as opposed to solely relying on subjective auditory impressions. Results demonstrated significant improvements resulting from visual feedback (de Bot, 1980) and crucially showed that a combination of visual and auditory feedback produced significantly more target-like pronunciation than auditory feedback alone (de Bot, 1983). While de Bot and colleagues focused more on the utterance-level intonation contours, more recent work suggests that visual feedback may also be relevant for intonation contours at the discourse level (Levis & Pickering, 2004).

Intonation represented an ideal speech feature for initial study, given both the general technological limitations of the time, as well as by the relative ease of interpreting visually presented intonation contours (Léon & Martin, 1972). However, with the more recent development of detailed, accurate, and accessible speech analysis software such as Praat (Boersma & Weenink, 2015), a number of researchers have begun to call for exploration of the potential use of such technology for instruction of segmental features (Chun, 2007). Broadly, software such as Praat allows for recording of speech stimuli and analysis via spectrograms and waveforms. While early work sought only to provide possible pedagogical implementations (Lambacher, 1999), others have attempted to examine the quantifiable benefits associated with visual feedback at the segmental level. Considering vowels, Saito (2007) demonstrated an increase in vowel pronunciation accuracy for Japanese-speaking learners of English following visual feedback and native model comparison, including both

spectrogram and waveform. Similar results have been found for vowel duration contrasts (Okuno, 2013). In contrast, other studies have found either no overall improvement in vowel segment accuracy, as was the case with English-speaking learners of French in Ruellot (2011), or mixed results depending on the vowel (Carey, 2004).

Fewer studies have considered the consonantal level, but results have been generally promising. Olson (2014b) examined the benefits of visual feedback and model comparison, in the form of both spectrograms and waveforms, on the production of voiced stops by English-speaking learners of Spanish. English is considered to have a single allophone for each of the voiced stops, whereas Spanish has two allophones in complementary distribution. Results demonstrated significantly more target-like productions of the voiced stops following the visual feedback paradigm relative to a control (non-treatment) group. Similar results have been found for geminate/singleton consonant production accuracy following a visual feedback paradigm (Motohashi-Saigo & Hardison, 2009). With respect to voice onset time, Lord (2005) showed significant improvements (i.e. more native-like) in voice onset time following a multi-faceted pronunciation training, including explicit phonetic instruction, transcription practice, oral practice, and visual feedback. However, as this study incorporated a multi-faceted pedagogical approach (Lord, 2005), the potential contribution of visual feedback on voice onset time remains unclear.

Taken as a whole, the research on visual feedback shows much promise, leading Motohashi-Saigo and Hardison (2009) to claim that "visual cues are a valuable source of input in L2 learning" (p. 42). If learners are not able to perceive their own (mis)pronunciation (see Dlaska & Krekler, 2008), then a paradigm such as visual feedback presents a way in which L2 learners can visually identify their errors and possibly achieve more success in self-correction. However, a number of clear areas for development remain. Given that such results have been most robust at the suprasegmental level, further research may continue to add to the mixed results at the segmental level, most specifically for consonantal segments. In addition, few

studies have looked specifically at the issue of the generalizability of pronunciation gains. Among the exceptions, Hardison (2004) showed that subjects were able to generalize gains made from utterance-level intonational training to novel utterances. From a perceptual approach, Motohashi-Saigo and Hardison (2009), revealed positive generalization to novel tokens following visual feedback and auditory training for the geminate/singleton consonantal distinction. However, in both cases, stimuli were limited to short, controlled, orthographically presented utterances. As such, it remains to be seen if gains made in production at the segmental level are generalizable, and specifically if there is generalizability from training in a restricted context (i.e. carrier utterance) to more cognitively challenging and naturalistic speech. This question is not only relevant at the pedagogical level, but may also have impacts at the theoretical level.

2.4. Research Questions

Drawing on previous research, this study investigates the use of visual feedback as a method of L2 pronunciation instruction for the segmental feature of voice onset time. Broadly, this work examines the effectiveness of the proposed visual feedback paradigm, as well as the generalizability of gains made from visual feedback on tokens in a restricted context to production in increasingly more unconstrained speech (e.g. novel utterances, continuous speech, and spontaneous speech). Specifically, this study seeks to investigate the following research questions: (1) Does a visual feedback paradigm improve production of the segmental feature voice onset time? (2) Do pronunciation gains in carrier utterances generalize to productions in novel utterances, continuous speech, and/or spontaneous speech?

3. Methods

In order to address these two research questions, this paper presents a visual feedback paradigm focused on target tokens with word initial voiceless stops. In addition, this study

examines recorded participant production across four different tasks (see section 3.2), from most restricted (i.e. carrier utterance) to controlled spontaneous speech.

3.1. Participants

A total of 24 participants, recruited from a large Midwestern university, were assigned to two groups: experimental group ($N = 17$) and control group ($N = 7$).[2] All participants were students in a fourth semester Spanish course (intermediate level). Drawing on a self-reported language background questionnaire, all subjects were native English speakers who had begun acquisition of Spanish after the age of 12, and who had minimal experience (less than 6 weeks) in a Spanish-dominant environment.[3] As participants likely had varied classroom experiences in the target language (e.g. number of years of instruction, native vs. non-native instructors, etc.), initial statistical analysis (see section 4) was employed to verify the homogeneity of the two groups.

Participants were required to complete the following tasks, detailed below, as part of the required coursework. The tasks were graded as a whole, in which the participants received full credit for simply having completed the task, not for how well any particular token was pronounced. After the final activity, participants were given the option, via informed consent, to provide their data for the current study. All participants, in both the experimental and control group, chose to provide their data for the research purposes. Participants received no compensation for their participation.

3.2. Stimuli

---

[2] As with most classroom-based research, "true" randomization of participant assignment to control or experimental groups was not feasible. Moreover, it is acknowledged that the sample size is relatively small. To address these issues, statistical analysis (below) demonstrates that both groups performed similarly on all pre-test measures.

[3] Five participants were eliminated from the original pool ($N = 34$) for failing to meet these criteria. Three were exposed to Spanish before the age of 12, one was a non-native speaker of English, and one had extended experience (greater than 6 weeks) in a Spanish-dominant context. An additional five participants were eliminated for failing to complete all parts of the study.

The current study focuses on the voiceless stops /p, t, k/ in word initial position. Voice onset time (VOT) can be defined as the beginning of vocal cord vibration following the release of the stop closure (Lisker & Abramson, 1964 among many). VOT durations in Spanish are considered to be short-lag (0-30ms) (e.g. Lisker & Abramson, 1964). In contrast, English VOT durations are considered to be long lag (50-120ms), although these values are dependent on stress, with word initial stops in unstressed position produced with somewhat shorter VOTs, albeit not within the Spanish range (Lisker & Abramson, 1967). This cross-linguistic difference in word-initial VOT often leads to native English learners of Spanish producing Spanish tokens with long-lag voiceless stops (Hammond, 2001; Hualde, 2005). With respect to phonological distribution, it is worth noting that English and Spanish broadly overlap in their implementation of word-medial VOT (i.e. short-lag) for voiceless stops (e.g. Hualde, 2005), and as such word-medial voiceless stops may not be problematic for non-native speakers.[4] Considering the impact of word-initial VOT on intelligibility and accentedness, Lord (2005) notes that English-like VOTs in Spanish tokens are "unlikely to cause confusion in meaning, they can result in a notable foreign accent" (p. 559).

The current study was comprised of four separate elicitation tasks, ranging from the most controlled (i.e. read-aloud carrier utterances) to the most spontaneous (i.e. picture naming task). These four tasks, detailed below, consisted of four different but related sets of materials.

3.2.1. Task 1: Carrier Utterance

Task 1 was comprised list of 30 target tokens within the carrier utterance *Di _____ de nuevo* ('say _____ again'). Target tokens consisted of Spanish words with the phonemes /p, t, k/ in initial position, followed by each of the 5 Spanish vowels /i, e, a, o u/ (3 stops x 5 vowels x 2 tokens = 30 tokens). Stimuli for Task 1 were produced by participants using a read aloud

---

[4] While there is some preliminary evidence that English word-medial VOT is increased in post-nasal position, it still falls within the short-lag VOT range (Hayes & Stivers, 2000). Further research may clarify the role of phonetic context and the potential perception of fine-grained VOT variability.

procedure. While each visual feedback paradigm focused only on one target phoneme, the same list of 30 tokens, including all three target stop consonants in carrier utterances, was used in each iteration of Task 1. Sample tokens are shown in Table 1.

Table 1. Sample target tokens in carrier phrase

Plosive	Example 1
/p/	Di *pesa* de nuevo
/t/	Di *testigo* de nuevo
/k/	Di *quepo* de nuevo

3.2.2. Task 2: Novel Utterances

Stimuli for Task 2 consisted of the same 30 tokens as in Task 1, but imbedded in novel utterances. During each iteration of Task 2, a subset of 15 target tokens were produced in novel utterances, which had never been seen by the participants. Stimuli for Task 2 were produced by participants using a read aloud procedure. Sample tokens for Task 2 are shown in Table 2.

Table 2. Sample target tokens in novel utterances

Plosive	Example 1	Example 2
/p/	Quiero un *poco* de agua.	No sé por qué *Paco* quiere irse.
/t/	Hay un *testigo* con el juez.	*Toca* la guitarra para mí.
/k/	Esa *cosa* no sirve para nada.	Llévame a *casa* por favor.

3.2.3. Task 3: Continuous Speech

Stimuli for Task 3 consisted of 30 target tokens, equally distributed among the three initial voiceless stop consonants /p, t, k/. Unlike Task 1 and Task 2, in which stimuli were presented in lists of simple, unrelated utterances, the target tokens in Task 3 were embedded within a short story (number of words = 479). Stimuli for Task 3 were produced by participants using a read aloud procedure. Thus, the controlled continuous speech task embedded tokens within a much broader context, and given the increased task load, directed attention and resources away from the individual targets. The same short story was used in each iteration of Task 3. Example 1 provides a short excerpt from the continuous speech task. Target tokens are underlined.

(1) *Me llamo __Paco__ y quiero contarte sobre mi primera experiencia __con__ mi compañero,*
Pedro. Había acabado de __cumplir__ 18 años, y tuve que mudarme a Indiana para mi
primer año de la universidad. Llegué a la casa de Pedro con mi __padre__ el 12 de
octubre… __Todo__ era perfecto, y en ese momento, no __quise__ ir adentro de la casa.

'My name is Paco and I want to tell you about my first experience with my roomate,
Pedro. I had just turned 18 years old, and I had to move to Indiana for my first year
of college. I got to Pedro's house with my father on the 12th of October. Everything
was perfect, and in that moment, I didn't want to enter the house.

3.2.4. Task 4: Spontaneous Speech Task

In order to create a context in which participants would produce the desired target tokens in a spontaneous or semi-spontaneous manner, a picture naming task was created based on that employed by Munro (2013). Stimuli for the picture naming task consisted of 30 pairs of visually presented pictures. In the first slide, participants were presented with a short utterance, including an orthographically presented target token, and accompanied by a picture of the target token. In the second slide, participants had to rephrase the utterance in the first slide, by both changing the name of the subject as well as inserting the correct word for the pictured, but not orthographically presented, target token. Thus, in observing the second slide, participants were required to produce both a new subject name as well as remember and produce the correct target token from the prior slide. Changing the name of the subject served as a distraction device, which did not permit participants to focus exclusively on the target token (Munro, 2013). Figure 1 represents an example of a pair of slides from the spontaneous speech task.

It should be recognized that the spontaneous task here is not indicative of truly "spontaneous" speech. The term *spontaneous* is used here to convey that, although it is still partially controlled, it represents more spontaneous production than the other three tasks included here.

Figure 1. Sample spontaneous speech task instructional slide (left) and elicitation slide (right) for the target token *pesa* 'weight'. Participants are directed to 'repeat the following sentence' and 'say the sentence with the picture.' A correct response would be *Pero ahora, Diego levanta la pesa,* 'But now, Diego lifts the weight.'

3.3. Procedures

Broadly, this study consisted of a pre-test, followed by three visual feedback interventions, and a post-test, conducted over the course of 8 weeks. Production of all stimuli were recorded by participants using Praat (Boersma & Weenink, 2015) at a 44.1kHz sampling rate and conducted in a quiet room. The order of the tasks and the visual feedback paradigm are described in below.

3.3.1. Task Order

The pre- and post-tests consisted of the recordings of each of the four tasks. The pre-test took place prior to any visual feedback paradigm, following a brief introduction to using Praat (Boersma & Weenink, 2015) to conduct audio recordings. The post-test took place two weeks after the final visual feedback paradigm.

For the experimental group, there were three visual feedback interventions (detailed below). The first visual feedback paradigm focused exclusively on training and improvement of the phoneme /p/. The second visual feedback paradigm focused on the phoneme /t/. And the final visual feedback paradigm focused on the phoneme /k/. Each of the visual feedback

paradigms included recordings of stimuli from Tasks 1 and 2 only. Actual visual analysis focused exclusively on tokens produced in Task 1, the carrier utterance task.

For the control group, recordings of the various tasks paralleled those done by the experimental group. The main difference between the two groups was that the control group received no explicit pronunciation training and did not participate in the visual feedback paradigm. Any pronunciation feedback that occurred during the training period for the control group consisted of the 'typical' ad-hoc approach to pronunciation previously reported via both observational (Foote et al., 2013) and self-reported methods (Foote, Holtby, & Derwing, 2011; Murphy, 2011, Olson, 2014b). Self-reporting by the instructor during a post-experiment debriefing confirmed that, in line with previous findings, the ad-hoc approach implemented throughout the normal course of instruction with both the control group and experimental group consisted of individual correction and pronunciation modeling. Pronunciation modeling includes, as classified by Lyster and Ranta (1997), *recasting* (i.e., reformulation of student production without the error) and *repetition* (i.e., repetition in isolation of the word or utterance containing the error). For both groups, no feedback of any kind was given on the target segments.

In addition, it is worth noting that the control and experimental groups received the same amount of instructional time in the classroom (120min/week). In order to ensure a similar amount of time dedicated to the target language, the control group spent a corresponding amount of time focused on socio-cultural aspects of the Spanish language relative to the time spent by the experimental group on the visual feedback paradigm. While this design was adopted for a more controlled experimental approach, there is a clear case for including both pronunciation and culture in regular language instruction (Kramsch, 1993). Both the control and experimental groups had the same, non-native, instructor.[5] While a degree of variation in

[5] As noted by one reviewer, non-native instructor modeling may impact student outcomes. However, as both groups had the same instructor, this impact should be similar in both groups.

instruction and interaction is inevitably present in this type of classroom environment, owing to individual student needs and group dynamics, every attempt was made to equate the experience of the two groups (e.g. same amount of time, same instructor, same coursework, etc.). Thus, while the experimental group received the visual feedback paradigm and the control group did not, the remainder of the course was similar. As such, the control group was chosen to represent a situation akin to what might be considered to be a common approach to pronunciation in the second language classroom.[6]

Table 3 illustrates the order of recording for both the control and experimental groups. Again, while only the experimental group performed the visual feedback paradigm, the control group performed the same recording tasks at the same specified times.

Table 3. Task Order

Session	Visual Feedback (Experimental Group)	Tasks Recorded
Pre-test		Tasks 1-4
Session 1	/p/	Tasks 1-2
Session 2	/t/	Tasks 1-2
Session 3	/k/	Tasks 1-2
Post-test		Tasks 1-4

3.3.2. Visual Feedback Paradigm

Each visual feedback paradigm consisted of an (a) initial self-recording, (b) guided visual analysis and comparison with a native speaker model, and (c) practice and re-recording (Olson, 2014a). Recordings were generally conducted at home and participants were simply instructed to record the stimuli in a quiet environment. This at-home recording methodology was chosen for both practical reasons (e.g., ease of implementation, limit use of class time) and to parallel previous research (Olson, 2014b).[7] For the initial self-recording, participants were

---

[6] Worth noting, this design addresses whether or not visual feedback is successful in pronunciation training, and specifically more successful than the "normal", minimal approach. However, whether visual feedback is more successful than other approaches to pronunciation instruction is left for future research.

[7] As noted by one anonymous reviewer, at-home recordings lack some of the control of laboratory-based recording procedures. To control for potential variation, explicit recording procedures were provided, in the target language, and no participants reported technical issues. Moreover, given a relatively small class size that allows for positive voice identification by the instructor and strict university-wide academic dishonesty guidelines, it is presumed that all participants created their own recordings.

provided with a set of stimuli, consisting of the stimuli in Tasks 1 and 2 (i.e. tokens in a carrier utterance and tokens in a novel utterance). Participants recorded the stimuli using Praat (Boersma & Weenink, 2015), printed the visual representation of five tokens, including both spectrograms and waveforms, for use in the visual analysis stage. Furthermore, participants were instructed to divide their words into the different 'sounds/letters', predominantly through repeatedly listening to the recorded word. Figure 2 illustrates a spectrogram produced by a native Spanish speaker (left) and native English speaker, respectively (right). Importantly, the visual feedback paradigm focused only on the stimuli in Task 1, therefore training was conducted only on the target words in the carrier phrases.
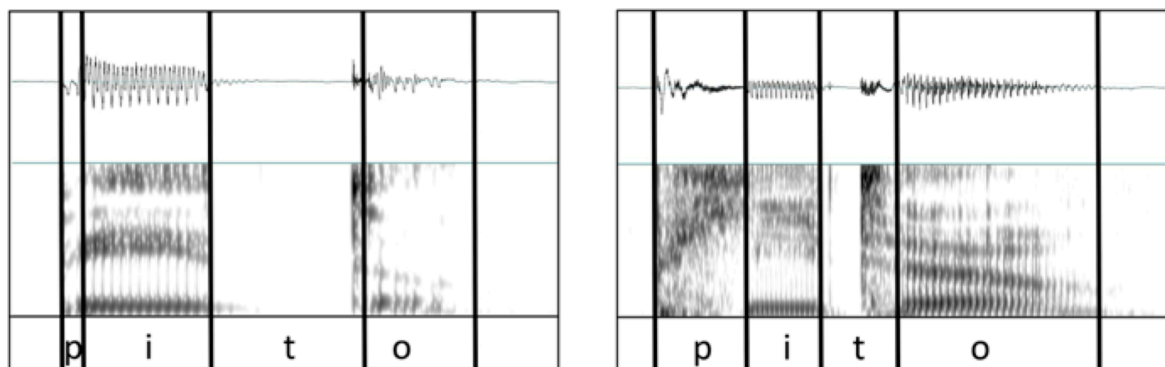

Figure 2. Sample waveform and spectrogram for the Spanish work *pito* 'whistle' produced by a native Spanish speaker (left) and a native English speaker (right). Voice onset time duration is indicated by the length of /p/.

For the visual analysis stage, participants were asked to answer a series of questions about their own productions, drawing focus to the pronunciation of the target phonemes (Example 2). All questions were provided in the target language.

(2)   How do you differentiate your vowel 'a' from the consonant 'p' in
       the word *Paco?*
       Is the 'a' longer or shorter than the 'p'?

Participants were then asked to answer a similar set of questions about the image of the same word produced by a native Spanish speaker, and compare/contrast their own productions with

those of the NS (Example 6). In short, participants were guided to describe the *visual* difference between their productions and those of a native Spanish speaker.

> (3)   What is the 'p' of the native Spanish speaker like?
> Have you noticed a difference in the length of the 'p' produced by the native Spanish speaker and your 'p'?

Following the visual comparison, participants heard the audio recording produced by the native Spanish speaker, and were asked to describe the *auditory* difference between their own productions and those of the native Spanish speaker.

Finally, following the above activity, participants were asked to re-record the same set of stimuli. Participants were allowed to re-record multiple times and were encouraged to examine the visual representations of their productions.

3.4. Analysis

Voice onset time (VOT) durations for each target token were analyzed and measured in Praat (Boersma & Weenink, 2015) with boundaries marked by hand, from the release of the stop consonant to the onset of voicing of the vowel (Lisker & Abramson, 1964), and durations were extracted using an automated script. Again, English has been shown to have long-lag VOT, while Spanish has short-lag VOT. As such, decreases in VOT imply more target-like production.

Statistical analysis was carried out using R v2.6.2 (R Development Core Team, 2008) using the LME4 package (Bates, Meachler, Bolker, & Walker, 2014). For all linear mixed effects models, the significance criterion was set at $|t| > 2.00$. The following subsections address the results for each of the four tasks.

4. Results

Results are presented below for each of the four tasks, from most restrictive to most naturalistic.

4.1. Task 1: Tokens in Carrier Utterance

A total of 3584 of a possible 3600 tokens (24 Participants × 30 Items × 5 Sessions [pre-test, Session 1, Session 2, Session 3, post-test] = 3600 tokens)[8] were included in the analysis of tokens in a carrier phrase. As can be seen in Figure 3 below, participants in the experimental group demonstrated general improvement over time with respect to VOT, from an average of 68.65ms (SD=27.70ms) in the pre-test to an average of 35.62ms (SD= 24.9ms) in the post-test. To determine if such improvement was significant, the data were submitted to an LME model with Group (experimental vs. control) and Time (pre-test, Phase 1, Phase 2, Phase 3, post-test) as fixed factors, and Participant and Item as random factors with both random slopes and intercepts (Barr et al., 2013).[9]

Of initial importance, results of the mixed model indicated no significant difference between the intercept (experimental group: pre-test) and the control group performance on the pre-test (β = -11.52, t = -.463), illustrating that both groups performed similarly prior to training. However, there was a significant interaction between the factors of Group and Time (at post-test: β = 34.57, t = 2.37), indicating that the control and experimental groups changed differently over the course of the study.

To better understand the data, subsequent mixed model analyses were conducted separately for the experimental and control groups, with fixed factor of Time and random factors of Participant and Item. Results for the experimental group demonstrate a significant difference between VOT at the pre-test (intercept) and each of the following sessions: Session 1 (β = - 23.62, t = -5.633), Session 2 (β = -30.85, t = -5.037), Session 3 (β = -28.28, t = -4.596), and post-test (β = -33.03, t = -5.416). Subsequent multiple comparison post-hoc analyses

---

[8] In each of the four tasks, a small number of tokens were eliminated due to speech errors (e.g. yawing, laughing), recording issues (i.e. background noise), or missing data (i.e. skipped by participant). Across all tasks, a total of 1.51% of the possible tokens were eliminated.

[9] Although /p, t, k/ have different VOT durations (e.g. Lisker & Abramson, 1964), the inclusion of *Item* as a random effect accounts for this inter-phoneme variation.

(TukeyHSD) demonstrated that while there was significant improvement between the pre-test and each of the other sessions, there were no other significant differences between any of the subsequent sessions (e.g. Session 1 vs. Session 2) (p > .1, for all subsequent comparisons). As such, the major gains are seen following the Treatment 1, as seen in Figure 3, and are maintained during all subsequent phases.

The performance of the control group stands out in contrast. Results of the mixed model analysis for the control group yielded no statistically significant differences for the tokens produced during the pre-test and any of the subsequent recordings ($|t| < .4$ in all cases). Visual analysis of Figure 4 illustrates this finding, with similar VOT durations maintained during each of the recording sessions.
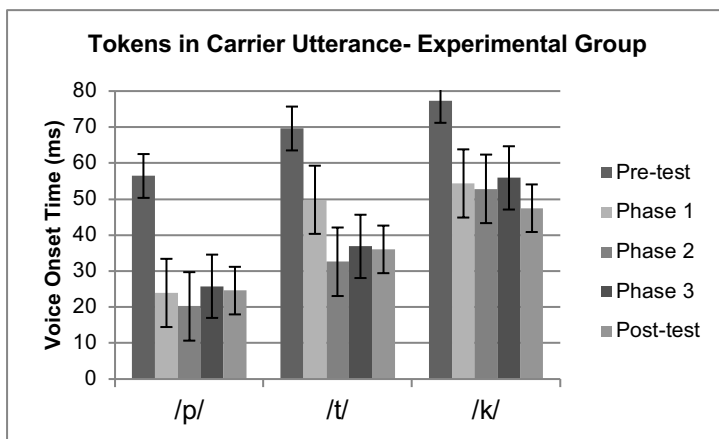

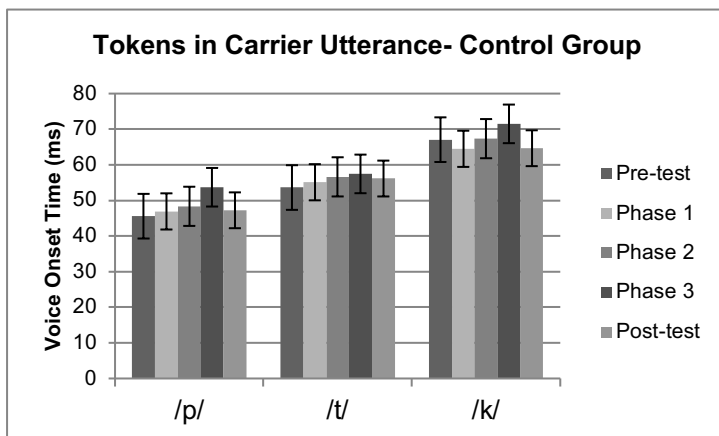Figure 3. VOT for target tokens in a carrier phrase produced by the experimental group.

Figure 4. VOT for target tokens in a carrier phrase produced by the control group.

4.2. Task 2: Tokens in Novel Utterances

When considering the performance of tokens within novel utterances, a pattern similar to that found for the tokens in carrier utterances emerges (Figures 5 and 6). A total of 1746 tokens were submitted to the statistical analysis (24 Participants × 15 Items × 5 Sessions = 1800 tokens). For the experimental group, the average VOT in the pre-test was 60.25ms (SD=31.1ms) and in the post-test 39.37ms (SD=26.66ms). For the control group, the average VOT for the pre-test was 52.62ms (SD=30.39ms) 49.6ms (SD=26.25ms) for the post-test. Statistical analysis, with a linear mixed model approach identical to that employed in the analysis of the tokens in carrier utterances, confirms the above observations.

Again, both groups were shown to perform similarly with respect to VOT duration in the pre-test, as illustrated by the lack of a significant difference between the intercept (experimental group: pre-test) and the control group performance during the pre-test ($\beta = -7.77$, $t = -0.590$). Subsequent models were conducted by group, with identical parameters as described above.

For the experimental group, significant reductions in VOT were found between the pre-test and: Session 1 ($\beta = -17.06.83$, $t = -4.795$), Session 2 ($\beta = -22.96$, $t = -2.612$), Session 3 ($\beta = -13.80$, $t = -2.188$) and the post-test ($\beta = -21.00$, $t = -3.426$). That is, the experimental group showed significant improvement following the first visual feedback paradigm and maintained such gains through the post-test.
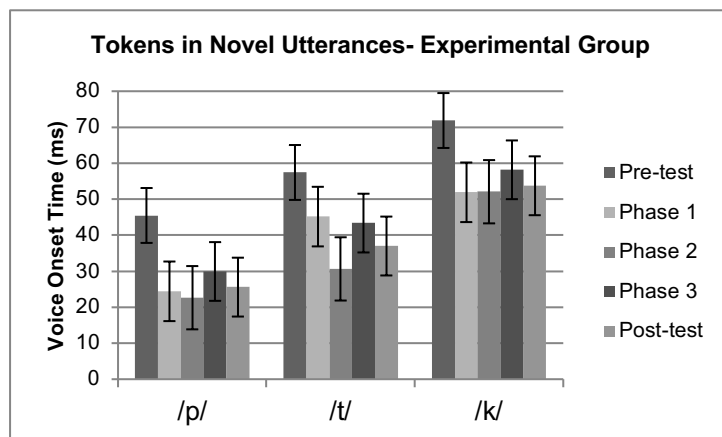

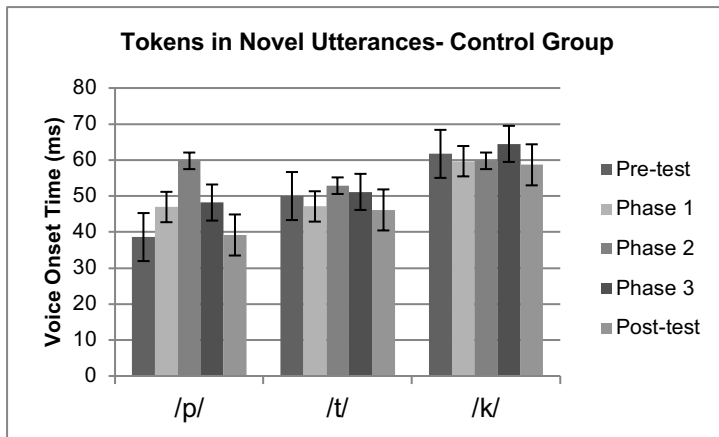Figure 5. VOT for target tokens in a novel utterance produced by the experimental group.

Figure 6. VOT for target tokens in a novel utterance produced by the control group.

For the control group, in contrast, there were no significant differences between the intercept (pre-test) and any of the subsequent recordings ($|t| < .5$ for all comparisons).

4.3 Task 3: Continuous Speech Task

A total of 1417 tokens were included in the analysis for the continuous speech task (24 Participants × 30 Items × 2 Sessions = 1440 tokens). It is worth repeating that the continuous speech task was performed only at the pre-test and post-test. The visual feedback paradigm never addressed tokens in continuous speech, and the continuous speech task was included as a measure of generalizability.

Results for the continuous speech task (Figure 7 and 8) parallel those found for the carrier utterance and novel utterance tasks. Specifically, while the experimental group produced longer aspiration values in the pre-test (M = 60.33ms, SD=30.77ms) than the post-test (M = 37.51ms, SD=26.6ms), the control group showed little variation between the pre-test (M=49.84ms; SD=26.26ms) and post-test (M=49.73ms; SD=27.06ms).

Statistical analysis confirms the significance of the above observation. Demonstrating the homogeneity of the two groups initially, there was no significant difference between the intercept (experimental group: pre-test) and the control group (pre-test) ($\beta = -10.42$, $t = -0.783$).

However, there was a significant interaction between the factors of Group and Time ($\beta = 21.67$, $t = 3.032$). Thus, while for the experimental group there was a significant difference between VOT production at the pre-test and post-test ($\beta = -22.67$, $t = -3.967$), there was no such difference for the control group. Again, these results indicate while both groups performed similarly initially, the experimental group showed significant improvement during the course of the training, while the control group showed no significant changes.
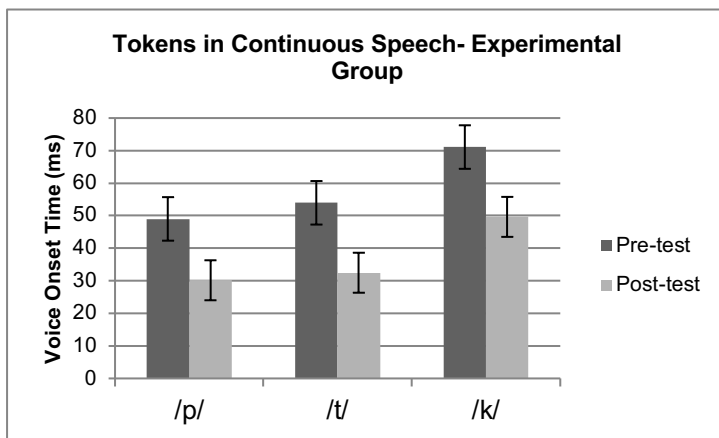

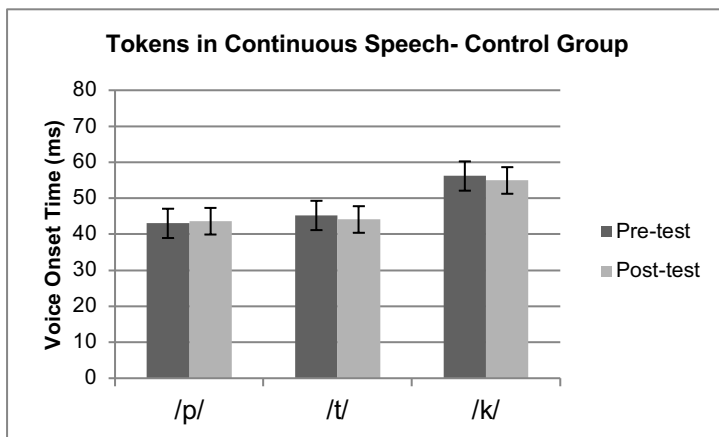Figure 7. VOT for target tokens in continuous speech produced by the experimental group.

Figure 8. VOT for target tokens in continuous speech produced by the control group.

4.4 Task 4: Spontaneous Speech Task

Results for the spontaneous speech task paralleled those found for the previous three tasks (Figures 9 and 10). As with the continuous speech task, the spontaneous speech task was performed only at the pre-test and post-test, and the visual feedback paradigm never addressed

tokens in spontaneous speech.

A total of 1408 tokens were included in the analysis of Task 4 (24 Participants × 30 Items × 2 Sessions = 1440 tokens), and a statistical approach identical to that used in section 4.3 was employed. While there was no significant difference between the two groups in the pre-test ($\beta = -14.62$, $t = -0.824$), there was a significant interaction between the factors of Group and Time ($\beta = 18.15$, $t = 2.826$). Again, the experimental group produced longer VOTs in the pre-test (M = 61.3ms, SD=30.22ms) than the post-test (M = 40.81ms, SD=24.82ms) ($\beta = -20.44$, $t = -5.169$), and the control group showed little difference between the pre- test (M=46.36ms; SD=22.02ms) and post-test (M=44.44ms; SD=21.49ms).
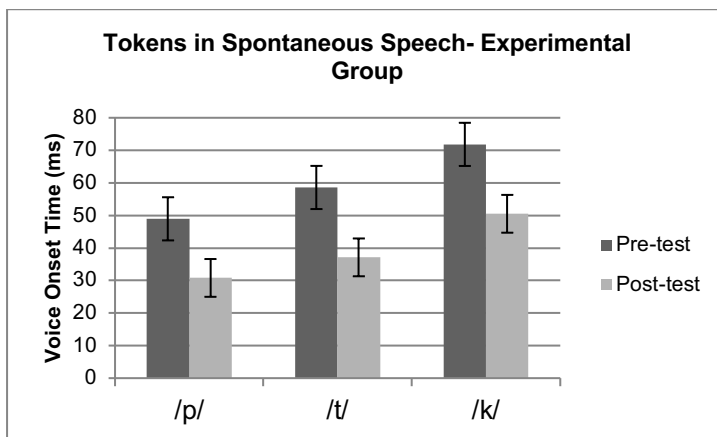

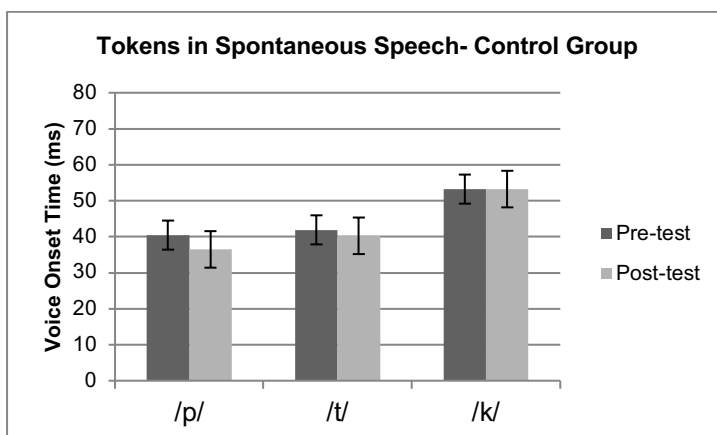Figure 9. VOT for target tokens in the spontaneous speech task produced by the experimental group.

Figure 10. VOT for target tokens in the spontaneous speech task produced by the control group.

4.5. Results by Phoneme

Although the above analysis showed a significant improvement concerning the experimental group in each of the four tasks, in order to better understand the effect of each visual feedback session, subsequent models and TukeyHSD multiple post-hoc comparisons were run for each of the individual phonemes produced in Tasks 1 and 2. It bears repeating that the phoneme /p/ was the focus of Session 1, /t/ the focus of Session 2, and /k/ the focus of Session 3.

With respect to the initial visual feedback paradigm (Session 1), which focused solely on /p/, reveals that there is a general trend towards improvement following Session 1 for all phonemes (/p, t, k/) produced in the carrier utterance task (Task 1). Subsequent statistical analysis (LMER) confirms this finding, with significant differences for each phoneme between the performance on the pre-test and following Session 1: /p/ (β = -32.53, t = - 4.848), /t/ (β = -18.01, t = -3.926) and /k/ (β = -21.26, t = -5.172). Parallel results were found for tokens in novel utterances (Task 2) (/p/ (β = -21.12, t = - 4.483), /t/ (β = -11.23, t = -2.184) and /k/ (β = -19.83, t = -4.246)).

Thus, while the visual feedback paradigm in Session 1 focused exclusively on the target phoneme /p/, significant improvement was evidenced for the non-target phonemes /t/ and /k/.

5. Discussion

The current study addresses the extant gaps in the literature, specifically related to the effectiveness of the visual feedback paradigm, particularly at the segmental level, as well as the potential generalizability of gains made during pronunciation instruction. Broadly, the results indicated that, following the implementation of a visual feedback paradigm, participants improved significantly in their productions of Spanish VOT. While tokens were not produced with strictly native-like VOTs, and this is rarely the individual or institutional aim for an L2 learner (e.g. Levis, 2007), there were significant gains made towards producing *more* native-

like productions of /p, t, k/. Furthermore, these gains were found not only in the restricted context that was the subject of the visual feedback paradigm, but these gains were found across the different tasks, from more constrained (i.e. carrier utterances) to least constrained (i.e. spontaneous speech).

5.1. Effectiveness of the Visual Feedback Paradigm on Segmental Instruction

Addressing the first research question, specifically to determine the effectiveness of the visual feedback paradigm, the current results add to a growing body of literature that support the efficacy of visual feedback for pronunciation instruction. While the earliest support for the use of visual feedback and modeling paradigms came from research at the suprasegmental level, specifically intonation (e.g. de Bot, 1980), the current study addresses previous calls in the literature for implementation of visual feedback at the segmental level (Chun, 2007). Whereas the limited existing research on visual feedback at the segmental level have focused predominantly on vowel segments (Carey, 2004; Okuno, 2013; Ruellot, 2011; Saito, 2007; although see Motohashi-Saigo & Hardison, 2009; Olson, 2014b), the current study shows the potential utility for visual feedback on the consonantal feature of voice onset time.

Moreover, while many of the previous studies have taken place either within a laboratory-based context (e.g. de Bot, 1980) or as part of a more advanced course focusing on L2 phonetics (e.g. Lord, 2005), this study is one of few to address the potential for pronunciation instruction at the lower levels of language instruction. The positive results seen here not only suggest that pronunciation instruction may be practical and relevant at the beginning stages of L2 instruction, but may serve to address the lack of empirically-tested pedagogical materials for pronunciation instruction. Important to note, the visual feedback paradigm employed here (see also Olson, 2014a) required no special technology during the class itself, assigned much of the pronunciation practice as homework, and incurred minimal

problems in its implementation (Olson, 2014a). While different feature and/or phonemes may require adaptation of the current paradigm, for example to account for phonemes with a complementary distributions, the results for word-initial VOT suggest that the visual feedback paradigm may be an effective method for pronunciation instruction.

In short, the visual feedback paradigm may be both successful and practical.

5.2. Generalizability of Pronunciation Gains

Although a number of studies have demonstrated significant improvement in L2 pronunciation owing to visual feedback, few have considered the issue of generalizability of pronunciation gains made in production. Generalizability, in this framework, can be defined as the transfer of pronunciation gains made during training to novel stimuli, and specifically novel stimuli in a more cognitively challenging and/or naturalistic context. At the suprasegmental level, Levis and Pickering (2004) considered the generalization of sentence-level intonation contours to new sentences. At the segmental level, Olson (2014b) investigated whether gains made during training on isolated words carried over to novel sentences, namely lists of numbered sentences that are simple in structure and do not pertain to one another. In the current paradigm, training occurred only on target tokens embedded in carrier utterance, however benefits extended to increasingly complex contexts, including novel utterances in list form, a long continuous speech sample (i.e. story), and a controlled spontaneous speech task.

These findings suggest generalizability is significant on several levels. On a practical and pedagogical level, they serve to validate the nature of gains made during the visual feedback paradigm. The visual feedback paradigm, as has been previously implemented, limits participants to examining tokens in controlled restricted contexts such as read-speech and short, simple utterances. For training to have real-world validity, such gains must necessarily apply beyond these restricted, unnatural contexts. At a theoretical level, while some authors have suggested that the effectiveness of the visual feedback paradigm lies in its ability to help learners "notice what they are doing" (Derwing & Munro, 2005, p. 387), the current results add

novel insight to our understanding of the reach of noticing. Bradlow and colleagues (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997) demonstrated a significant improvement in production driven by perceptual training, suggesting that noticing at the perceptual level has ramifications in production. The current work builds on this notion of noticing, illustrating that noticing in a restricted context may have ramifications in more complex, naturalistic contexts.

As a secondary note, while generalizability was originally defined in this study to refer to the carry-over of gains in pronunciation made following training on a carrier utterance to more spontaneous speech, a degree of phoneme-level generalizability was also found. That is, following explicit visual feedback on the /p/ segment in Session 1, participants significantly improved pronunciation of the segments /t/ and /k/. While not the focus of the current study, these results are promising. From a pedagogical stand-point, understanding the generalizability of certain features may serve to make pronunciation training more efficient. That is, training one feature may improve pronunciation of multiple phonemes. As a corollary from a more theoretical framework, although outside the scope of the current work, the interplay of gains made on differing phonemes may add to our understanding of the development of the L2 phonemic system.

6. Conclusion

The current paper addresses the general lack of research on pronunciation instruction, specifically by investigating the use and benefits of a visual feedback paradigm at the segmental level. Two important findings emerged from the results. First, the current results demonstrate the viability of the visual feedback paradigm on a previously unattested segmental feature, namely voice onset time. Taken in conjunction with previous research regarding visual feedback on other segmental features, most commonly vowels, the current findings add support for the use of visual feedback in the classroom for a variety of phonetic features. Second, and perhaps more importantly, the results of the current study suggest a degree of generalizability

for gains made using the visual feedback paradigm. Specifically, while training addressed only tokens produced in a carrier sentence, the pronunciation gains made via the visual feedback paradigm extended to tokens in novel sentences, tokens in continuous speech, and tokens in spontaneous speech.

While the current study presents only one possible iteration of a visual feedback paradigm, future research and technological development will only serve to enhance the current activities. Research may focus on identifying other sounds that lend themselves to this, and other visualization techniques. As suggested by Olson (2014b), some phonetic features, such as duration, intensity, and intonation, may be inherently more intuitive for visual inspection than others (e.g. rhoticity, nasality). Furthermore, other types of visualization, including transient feedback (e.g. Hincks & Edlund, 2009), electropalatography, or ultrasound, may lend themselves best to different features. It should also be acknowledged that the current findings are drawn from a relatively small sample size and focused on a single phonetic feature. As such, both students and researchers would be well served by replication with larger and more varied groups of learners, as well as other phonetic features. In addition, while the current study design sought to examine the usefulness of the visual feedback paradigm as compared to the relatively minimal "traditional" (e.g. ad hoc) approach to pronunciation, it remains to be seen if the gains found here are any different from other types of pedagogical approaches (e.g. explicit articulatory instruction). Future research may seek to compare multiple types of pronunciation trainings to determine the relative usefulness of different approaches. In short, while the visual feedback paradigm shows much promise, both in terms of utility and generalizability, much work remains to be done.

References

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

Anderson-Hseih, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5–22.

Arteaga, D. (2000). Articulatory phonetics in the first-year classroom. *The Modern Language Journal, 84,* 339–354.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255-278.

Bates D, Maechler M, Bolker B and Walker S (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Blas Arroyo, J. L. (1999). Las actidudes hacia la variación interdialectal en la sociolinguistica hispánica. *Estudios Filológicos, 34*, 47–72.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (version 5.4.08) [computer software]. Available from 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.

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.

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

21, 46-66.

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., 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. (1997). Accent, intelligibility, and comprehensibility: Evidence

from four L1s. *Studies in Second Language Acquisition, 20*, 1–16.

Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching: A

research-based approach. *TESOL Quarterly, 39*, 379–397.

Derwing, T., & Munro, M. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476–490

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.

Dlaska, A., & Krekler, C. (2008). Self-assessment of pronunciation. *System, 36*, 506-516.

Elliott, R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal, 79*(4), 530–542.

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

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.

Gluszek, A., & Dovidio, J. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in Communication. *Personality and Social Psychology Review, 14*(2), 214–237

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.

Hayes, B. & Stivers, T. (2000). *Postnasal voicing*. Unpublished manuscript, University of California–Los Angeles, Los Angeles, CA.

Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentation with transient visual feedback. *Language Learning and Technology, 13*(3), 32-50.

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

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal, 27*(1), 1-12.

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

Kim, S. Y., Wang, Y., Deng, S., Alvarez, R., & Li, J. (2011). Accent, perpetual foreigner stereotype, and perceived discrimination as indirect links between English proficiency and depressive symptoms in Chinese American adolescents. *Developmental Psychology, 47*(1), 289–301.

Kramsch, C. (1993). *Content and culture in language teaching.* Oxford: Oxford University Press.

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.

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

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

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.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical Measurements. *Word, 20*(3), 384-422.

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

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.

Macdonal, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation. The variable effects of different types of instruction. *Language Learning, 44*(1), 75-100.

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.

Munro, M. (2013). *What do you know when you "know" an L2 vowel?* Paper presented at Pronunciation in Second Language Learning and Teaching conference, Ames, IA.

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

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

Munro, M., Derwing, T., & Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language Teachers. *Prospect, 21*(1), 67-79.

Murphy, D. (2011). An investigation of English pronunciation teaching in Ireland. *English Today, 27*(4), 10–18.

Neufeld, G. (1978). On the acquisition of prosodic and articulatory features in adult languaeg learning. *Canadian Modern Language Review, 34,* 161-174.

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. (2014a). Phonetics and technology in the classroom: a practical approach to using speech analysis software in L2 pronunciation instruction. Hispania, 97(1), 1–25.

Olson, D. (2014b). Benefits of visual feedback on segmental production in the L2 classroom. Language Learning and Technology, 18(3), 173–192.

Pennington, M., & Ellis, N. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*(3), 372–389.

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

Purkiss, S., Perrewé, P., Gillespie, T., Mayes, B., & Ferris, G. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes, 101*, 152–167.

Rosenman, A. (1987). The relationship between auditory discrimination and oral production of Spanish sounds in children and adults. *Journal of Psycholinguistic Research, 16*(6), 512–34.

Rossiter, M., Derwing, T., Manimtim, L., & Thomson, R. (2010). Oral Fluency: The neglected component in the communicative language classroom. *The Canadian Modern Language Review, 66*(4), 583–606.

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.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Saalfeld, A. (2011). *Acquisition of L2 phonology in advanced learners: Does instruction make a difference?* In. J. Levis & K. LeVelle (Eds.). Proceedings of the 2nd Pronunciation

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

Saito, K. (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.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. *Language Awareness, 20*(1), 45-59.

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

Saito, K. (2013). Reexamining effects of Form-Focused Instruction on L2 pronunciation development. *Studies in Second Language Acquisition, 35*, 1-29.

Santos Maldonado, C. (1994). Some aspects of 'foreignness' in the pronunciation of upper intermediate English students of Spanish. In A. Davies and B. Parkinson (Eds.) *Edinburgh working papers in applied linguistics, 5,* (pp. 78-97)*.* Edinburgh: Edinburgh U. 78–97.

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.

Sturm, J. (2013). Explicit Phonetics Instruction in L2 French: A Global Analysis of Improvement. *System, 41*, 654-662.

Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning, 26*, 233–253.

Thomson, R., & Derwing, T. (2004). Presenting Canadian values in LINC: The roles of textbooks and teachers. *TESL Canada Journal, 21*(2), 17-33.

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.