VISUAL FEEDBACK AND RELATIVE VOWEL DURATION IN L2 PRONUNCIATION: THE CURIOUS CASE OF STRESSED AND UNSTRESSED VOWELS

Daniel J. Olson, Purdue University

Visual feedback for pronunciation training consists of providing learners with visual representations of their own productions and facilitating comparison with a native speaker model. While visual feedback has been shown to be successful in training some absolute duration-based consonantal cues (e.g., voice onset time), the current study examines whether visual feedback may improve relative durational contrasts (e.g., stressed vs unstressed vowels). Using a pretest, intervention, and posttest design, intermediate-level English-speaking learners of Spanish completed a visual feedback paradigm focused on relative vowel duration. In English, stressed vowels are significantly longer than unstressed vowels (i.e., twice as long). In Spanish, stressed vowels are only marginally longer than unstressed vowels. The intervention consisted of three activities in which participants read aloud and recorded utterances containing target words and compared their spectrograms/waveforms with those produced by native speakers. Target tokens were controlled for phonetic environment, syllable structure, and cognate status. Contrary to original hypotheses, results showed that at the pretest, learners produced shorter stressed vowels and longer unstressed vowels, a pattern not observed in either English or Spanish. Furthermore, no change was found following the visual feedback paradigm. The results are discussed with reference to cognitive load and methodological trade-offs between controlled and spontaneous speech.

Cite as: Olson, D. J. (2022). Visual feedback and relative vowel duration in L2 pronunciation: the curious case of stressed and unstressed vowels. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13353

INTRODUCTION

Visual Feedback

In recent years, visual feedback has emerged as a viable method for teaching L2 pronunciation. Broadly, visual feedback consists of providing learners with visual representations of their speech, or some aspect of their speech, often accompanied by a visual representation of native speaker productions. Direct visual feedback is used to describe paradigms that directly show articulatory movements during speech (for indirect vs. direct see Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015). Indirect visual feedback relies on the illustration of some acoustic speech property, such as intonation contours (Hardison, 2004), spectrograms (Saito, 2007), and/or waveforms (Olson, 2014). This feedback can be given in real-time (Garcia, Kolat, & Morgan, 2018), but more commonly is presented immediately following the learner production. Learners must then link these abstract representations to articulatory movements.

While early visual feedback paradigms focused on suprasegmental features, notably intonation contours (e.g., Anderson-Hseih, 1992; de Bot, 1980; Chun, 1998), more recent work has begun to leverage visual feedback for teaching segmental features, such as consonants and vowels. A range of features have been shown to improve following visual feedback, including voice onset time (Offerman, 2020; Olson, 2019), singleton/geminate contrasts (Motohashi-Siago & Hardison, 2009), vowel duration (Okuno, 2013), and intervocalic spirantization (Olson, 2014). Importantly, gains made following visual feedback have been shown to generalize to non-trained stimuli (Hardison, 2004; Offerman, 2020) and appear to be long-lasting (e.g., Olson & Offerman, 2020). Moreover, recent studies have shown that visual feedback can be incorporated in intact classes at the beginning and intermediate levels.

Visual feedback may be successful as it enhances L2 learners' abilities to notice differences between their speech and that of a native speaker. This additional noticing may serve to enhance learners' understanding of acoustic differences that they have already successfully perceived, or provide a new modality to allow them to notice a difference that they were previously unable to perceive. The Noticing Hypothesis (Schmidt, 1990) suggests that awareness of L2 forms, and specifically differences between the L1 and L2, is a necessary precondition for some types of acquisition. There has been further suggestion that noticing may be important for acquiring L2 pronunciation (Derwing & Munro, 2005), as L2 phonemes that are similar to existing L1 phonemes may be subsumed by the L1 category, preventing establishment of a new L2 category (for example, Perceptual Assimilation Model- L2 (PAM-L2): Best & Tyler, 2007). Furthermore, linked directly with the Noticing Hypothesis, visual feedback may be considered relative to the framework on corrective feedback, which has been shown to be most effective when learners are provided model pronunciation forms (for review of corrective feedback in L2 pronunciation see Saito, in press).

Drawing on the notion of noticing, Olson (2014) has argued that the success of visual feedback for a given speech feature may be limited by how intuitive the visual representation is to L2 learners. The features that have been most successfully improved with visual feedback at the suprasegmental and segmental levels, intonation and consonant duration respectively, may be both easy to perceive visually and easy to link to the corresponding articulately movement. Other features, such as the location of an L2 vowel in vowel space, which has less consistently evidenced improvement following visual feedback (Carey, 2004; Ruellot, 2011), may be more difficult to perceive and/or link to the corresponding articulatory gestures.

While consonant duration has been shown to be consistently improved, several different types of durational features are relevant in phonetic production. Specifically, while some consonants rely on absolute durations (e.g., voice onset time), other features rely on relative duration measures (e.g., stressed and unstressed vowels). For example, in English, voiceless stop consonants are generally produced with a voice onset time between 30–100ms (Lisker & Abramson, 1964). In contrast, to determine whether a given vowel or syllable is stressed, listeners must compare durations across multiple vowels or syllables. In English, listeners are unable to determine if a vowel with a duration of 200ms is stressed or unstressed, without comparing it to surrounding vowels (Ortega-Llebaria, Olson, & Tuninetti, 2018) As such, it remains to be seen whether visual feedback is an effective method of teaching such relative durational contrasts.

Vowel Duration in English and Spanish

English and Spanish differ with respect to the relative durational contrast involved in distinguishing stressed and unstressed vowels (Figure 1). In English, stressed vowels are significantly longer than unstressed vowels, with some research showing that stressed vowels are 2.2 times longer than unstressed vowels (de Jong, 2004). In addition, unstressed vowels undergo vowel reduction, including both a centralization in the vowel space and a shortening of duration. In Spanish, stressed vowels are only marginally longer than unstressed vowels (1.1 times longer; Nadeu Rota, 2013), and there is no systematic vowel reduction in most dialects. Given this cross-linguistic difference, some authors have noted that English-speaking learners of Spanish often implement their English duration norms when speaking Spanish (Hammond, 2001). English speaking learners of Spanish may produce longer-than-expected stressed vowels and/or shorter-than-expected unstressed vowels, giving their pronunciation an "inappropriate" or "sing-song" quality (Hammond, 2001, p. 314). This cross-linguistic transfer likely results in issues of comprehensibility and accentedness, and potentially issues in intelligibility (for differentiation of intelligibility, comprehensibility, and accentedness, see Munro & Derwing, 1995), as stress is contrastive at the lexical level in Spanish (e.g., hablo ['a.blo] 'I speak' vs. habló [a.'blo] 'he spoke').

Research Question

In light of previous research that has shown that visual feedback is successful for training *absolute* L2 consonant durational features, the current project examined the efficacy of visual feedback for training *relative* durational features. Specifically, the guiding research question was: Does visual feedback, in the form of waveforms and spectrograms, serve to improve L2 production of relative vowel duration?

Drawing on previous research, there were two initial hypotheses. First, it was anticipated that L1 English learners of L2 Spanish would demonstrate a degree of L1 transfer of relative vowel duration. It was expected that stressed vowels would be significantly longer (twice as long) than unstressed vowels. This first hypothesis essentially serves as a necessary precondition for examining the impact of visual feedback. Second, it was hypothesized that learner productions of relative vowel duration would become more native-like following visual feedback training. Specifically, it was expected that the difference in stressed and unstressed vowel durations would diminish following the visual feedback activities.

METHODS

Participants

Thirteen participants were initially recruited from an intact intermediate-level Spanish language classroom at a large public university. This course, the third in a sequence of six Spanish language courses, focused on the four main language skills (reading, writing, speaking and listening), and as well as elements of Hispanic culture. Spanish courses focused on A total of 6 participants were retained for the final analysis, with 6 eliminated for failing to provide either pretest or posttest data and one for failing to meet the inclusionary criteria (native speaker of English). Participants provided background information via an abbreviated version of the self-reported Bilingual

Language Profile (BLP: Birdsong, Gertken, & Amengual, 2012). All participants were native speakers of English, acquiring English from birth (Mean AoA = 0.0, *SD* = 0.0) and Spanish after the age of 7 (Mean AoA = 13.0, *SD* = 3.0). All participants are considered to be English-dominant (see Table 1).

Table 1

Participant Profile

Subcomponent	Measure	English *M* (*SD*)	Spanish *M* (*SD*)
Language Acquisition	Age of Acquisition (years)	0.0 (0.0)	13.0 (3.0)
Language Use	Percentage of Use (0–100)	92.2 (11.4)	4.5 (4.3)
Language Proficiency	Likert Scale 1–7[a]	7.0 (0.0)	3.7 (0.8)
Language Attitudes	Likert Scale 1–7[b]	6.5 (0.5)	2.8 (1.0)

[a] *1 = do not speak language at all; 7 = speak language very well*

[b] *1 = do not identify with culture at all; 7 = identify with culture very strongly*

Stimuli

Stimuli consisted of 72 two-syllable paraoxytonic words (36 pretest; 36 posttest). All target words were included in utterance-initial position within unique utterances. All words were non-cognate. As previous research has shown that different vowels have inherently different durations (Toivonen et al., 2015), stimuli were balanced for stressed vowel (/u, o, a/). Due to the limited number of real-word stimuli fitting the strict controls, 12 tokens (16.7%) included diphthongs in the stressed vowel position, potentially artificially increasing the stressed vowel duration and vowel duration ratio. Given the results, this is unlikely to have significantly impacted the outcome of the study. Finally, as word frequency has been shown to impact phonetic production (Jurafsky, Bell, Gregory, & Raymond., 2001), tokens were balanced for frequency across the two sessions.

Word familiarity was considered a more appropriate measure for L2 learners than word frequency (Auer, Bernstein, & Tucker, 2000). To control for word familiarity, a second set of L1 English – L2 Spanish speakers were recruited for familiarity norming (N = 21). Participants were given a randomized list of the target words and asked to rate them on a familiarity scale: 1 = highly unfamiliar; 7 = highly familiar (Auer et al., 2000). Results of a non-paired t-test showed no significant difference in the familiarity of target tokens between the pretest (M = 4.5, SD = 2.3) and posttest (M = 4.7, SD = 2.1), $t(69.4) = -0.240$, $p = .811$. Table 2 provides sample stimuli.

Table 2

Sample Stimuli and Corresponding Translations

Sample Stimuli	English Translation
Tuve una cafetera, pero se me rompió.	I had a coffee maker, but it broke.
Poda las ramas de este árbol.	Trim the branches of this tree.
Tacha mi nombre de la lista.	Cross my name off the list.

Procedures

This study employed a pretest, intervention (×3), and posttest design. The pretest was conducted immediately prior (< 3 days) to the first visual feedback treatment. The posttest was conducted immediately following (< 3 days) the final visual feedback treatment. Pre- and posttest stimuli were presented in a written list and were recorded using Praat (Boersma & Weenink, 2018) on the participants' home computers using a microphone. Although students were instructed to record the stimuli in a quiet location, given that duration measurements are robust against poor recording quality effects, no additional efforts were made to control recording equipment or environment.

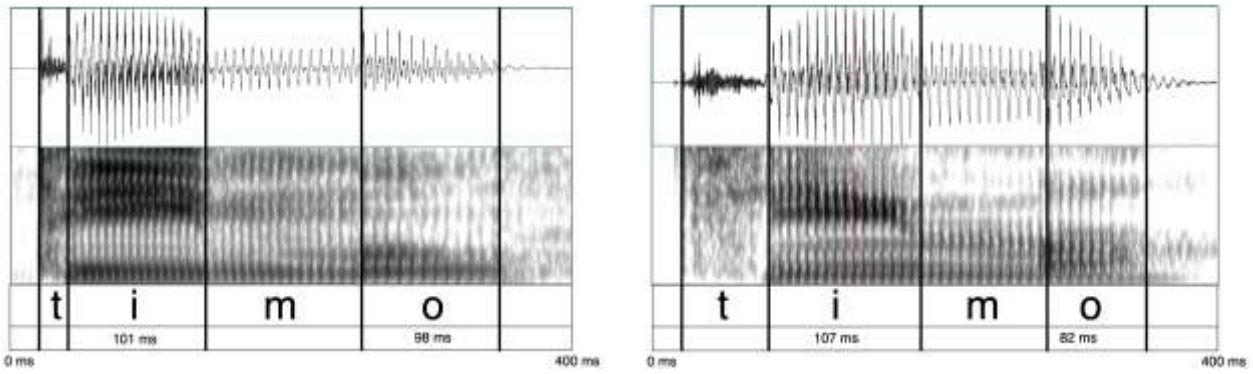
Each visual feedback activity included: (a) at-home recordings, (b) guided in-class visual analysis of learner productions, (c) guided in-class visual analysis of native speaker productions, and (d) at-home re-recording. During the in-class phase, which lasted approximately 25 minutes per session (× 3 sessions), participants analyzed their own productions (printed, using Praat) and native speaker productions in small groups following prepared guiding questions in the target language (Example 1).

(1) In small groups, look at the images of the word *pido* that you recorded and answer the following questions:
   a. How did you decide to mark the boundaries of each sound?
   b. What are the visual characteristics of the "i" and the "o"?
   c. Is your "i" long or short compared with your "o"?

After analyzing their own productions, participants compared their waveforms with those of a native speaker and were asked to hypothesize about the auditory differences. To emphasize the difference, participants compared several pairs of waveforms and spectrograms produced by native and non-native speakers (see Figure 1). Finally, participants were asked to record a new set of stimuli at home within 3 days of completing the in-class activity. Participants were encouraged to record multiple times and examine their waveforms and spectrograms.

Figure 1

Waveform and spectrogram of the word timo ['ti.mo] 'scam' produced by a native Spanish speaker (left) and a native English speaker (right).



While each visual feedback activity focused on the same feature (stressed and unstressed vowel duration), they increased in complexity. Visual feedback activity 1 focused on words in isolation, activity 2 focused on words in utterances, and activity 3 focused on words in a connected paragraph. Tokens examined during the in-class phase were different from those used for analysis.

The visual feedback activities, including the pretest and posttest recordings, formed part of the course curriculum. Grades were given on a complete/incomplete basis, and no additional feedback was given to learners. Following the completion of the curricular activities, participants were invited to provide their data for the research project. Those that consented filled out the language background questionnaire and were compensated for their participation.

Analysis

A total of 432 tokens were included in the initial data analysis (6 participants × 36 tokens × 2 sessions). Participant productions were measured by a single, trained coder using Praat (Boersma & Weenink, 2018). Approximately 7.6% of tokens were eliminated from the analysis (production errors n =15; background noise n = 2); outliers (vowel duration ratio ± 2 SD; n = 16)), leaving 399 tokens in the final analysis.

A *vowel duration ratio* was calculated by dividing the duration (ms) of stressed vowels by the duration (ms) of unstressed vowels. A ratio greater than 1 corresponds to stressed vowels that are longer than unstressed vowels. A ratio less than 1 corresponds to stressed vowels that are shorter than unstressed vowels. Based on previous research, the expected vowel duration ratio for English (or English-accented Spanish) is approximately 2.2 and Spanish is approximately 1.1.

Linear mixed effects models were conducted using R (R Core Team, 2013) and the lme4 package (Bates et al., 2014), using a maximal random effects structure (Barr et al., 2013). The significance criterion was set at $|t| = 2.00$.

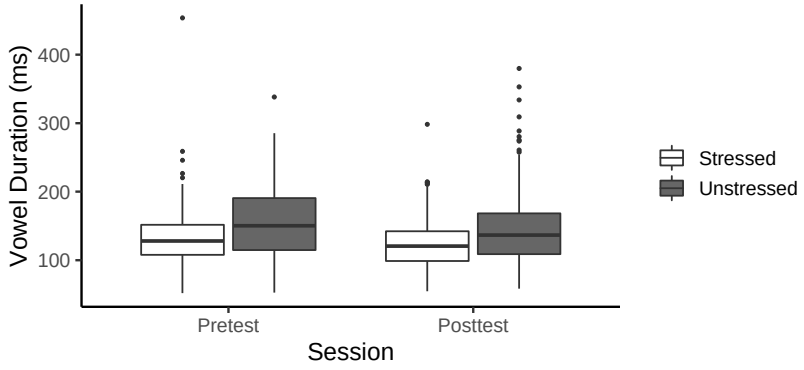
RESULTS

A mixed effects model was conducted with raw vowel duration (ms) as the dependent variable, vowel (stressed vs. unstressed) and session (pretest vs. posttest) as fixed effects, and participant and tokens as random effects. The maximal random effect structure that permitted model convergence was random intercepts for both participant and token, and random slope by session for participant.

Results of this initial model indicated a significant difference between the intercept (pretest, stressed vowels) and the pretest unstressed vowels (b = 23.114, SE = 3.789, t = 6.100). However, an examination of the direction of this difference (Figure 2) shows that, contrary to expectations, the stressed vowels were actually shorter than the unstressed vowels. This was found in both the pretest (stressed M = 133.2 ms, SD = 40.0; unstressed M = 156.3 ms, SD = 52.2) and the posttest (stressed M = 124.0 ms, SD =35.6 ; unstressed M = 146.4 ms, SD = 54.6). There was no difference between the intercept and the posttest (b = –8.404, SE = 10.663, t = –0.788), and no interaction between stress and session (b = –0.768, SE = 5.463, t = –0.141).

Figure 2

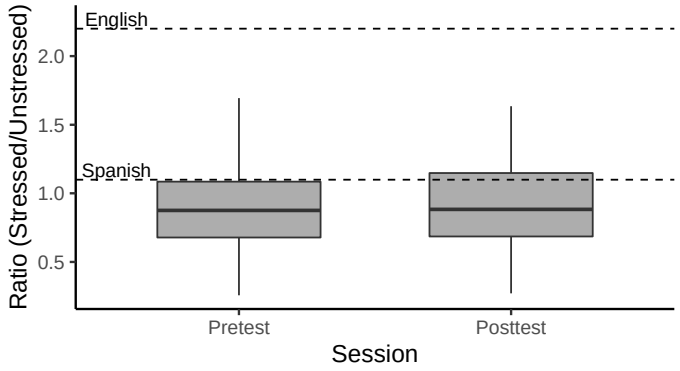
Vowel duration (ms) by stress and session.



Assessing duration as a relative measure, a second set of analyses was conducted on vowel duration ratio. The mixed effects model included vowel duration ratio as the dependent variable, session as the fixed effect, and participant as a random effect, with random intercept and slope (by participant). Results (Figure 3) revealed no significant difference between the pretest and posttest (b = 0.001, SE = 0.032, t = 0.031), showing that the visual feedback activities had no impact on the overall duration contrast employed by participants. Worth noting, the overall mean vowel duration ratio (M = 0.921, SD = 0.311) was well below the expected (transferred) English value of approximately 2.2 (de Jong, 2004). However, participants were not performing like native Spanish speakers, as they produced vowel duration ratios that were below even the expected Spanish norms (Nadeu Rota, 2013).

Figure 3

Vowel duration ratio by session. The dashed horizontal lines provide expected norms for English (de Jong, 2004) and Spanish (Nadeu Rota, 2013).



Given the unexpected results, a final layer of analysis was conducted using only a subset of the tokens representing the most controlled syllable structure (CVCV). This analysis was conducted to ensure that the results were not influenced by that variable syllable structure found in the stimuli. In total, 22 tokens consisting of only CVCV syllables were included in this final analysis (pretest

= 13, posttest = 9), resulting in 126 vowel duration ratio measurements. Model parameters were simplified to permit convergence, with participant as a random effect with random intercept only. The results closely parallel those found for the whole dataset. Specifically, there was no difference between the vowel duration ratio at the pretest (M = 0.880, SD = 0.273) and the posttest (M = 0.964, SD = 0.371) (b = 0.078, SE = 0.055, t = 1.413). Again, vowel duration ratios below 1.0 suggest unstressed vowels that are *longer* than stressed vowels, a pattern not expected in either the L1 (English) or L2 (Spanish).

DISCUSSION

Taken as a whole, the original research question could not be answered by the current results. Specifically, as learners did not initially (i.e., at the pretest) produce stressed and unstressed vowel duration with English-like durations (i.e., Hypothesis 1 not substantiated), it was not possible to evaluate the impact of the visual feedback training.

While there are several possible explanations for the current results, one in particular should be noted. The current study employed a read-speech paradigm, in which intermediate-level learners are asked to read in the target language. Relative to spontaneous speech, the additional cognitive load of the reading task may impact speech production. Previous research with monolinguals has suggested that an increase in cognitive load results in greater variability in articulation rate, an increase in "drawls", in which speakers elongate syllables during hesitation, and an increase "in major prosodic boundaries in minor syntactic units" (Christodoulides, p. xiv). In English, both drawls and major prosodic boundaries result in lengthening of the unit-final syllable (Klatt, 1973). In the current stimuli, this lengthening, resulting from the increased cognitive load and use of hesitation-oriented lengthening or inclusion of a prosodic boundary would apply to the word-final, unstressed vowel. In short, lengthening would disproportionately impact the unstressed vowel, resulting in an artificially lowered vowel duration ratio, as was seen in the current data. While this explanation seems reasonable, future research would be needed to consider this possibility.

In light of the unexpected results, it is worth considering the implications for researchers in the field of second language pronunciation. Considering differing methodologies in second language pronunciation research, a number of authors have called for more use of spontaneous speech data in L2 pronunciation research (e.g., Thomson & Derwing, 2015). The current study, in which results may have been significantly impacted by the reading paradigm, coupled with previous research that successfully employed similar methodologies for other features (for VOT see Offerman & Olson, 2016), suggest a nuanced, feature-specific approach to study procedures. In contrast, other features, such as relative vowel duration, may be impacted more by controlled paradigms, and may benefit from a more spontaneous speech approach. Moreover, differing methodologies fall along a continuum of ecological validity, from the most controlled elicitations (e.g., words in isolation, carrier phrases) as the least ecologically valid, to the most spontaneous, naturalistic conversation as the most ecologically valid. The method used here, unique utterances read aloud, may represent a middle-ground on this continuum. Given the inherent trade-offs between controlled (more control over phonetic environment, more repetitions of target features and potentially larger data sets, less natural) and spontaneous speech (less control over phonetic features, fewer repetitions, more natural), researchers may choose to tailor and pilot their experimental approach to the feature under examination.

ABOUT THE AUTHOR

Dr. Daniel J. Olson (danielolson@purdue.edu) is Associate Professor of Spanish and Linguistics at Purdue University and Director of the Purdue Bilingualism Lab. His research focuses on phonetics and psycholinguistics related to bilingual populations. He is particularly interested in the cognitive mechanisms and pedagogical methods that facilitate the acquisition of second language phonetics. His work also examines the production and perception of code-switching, as well as the underlying mechanisms that govern bilingual language separation and selection.

Purdue University
School of Languages and Cultures
640 Oval Dr.
West Lafayette, IN 47906

REFERENCES

Anderson-Hseih, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20,* 51–62. https://doi.org/10.1016/0346-251X(92)90007-P

Auer, E. T., Bernstein, L. E., & Tucker, P. E. (2000). Is subjective word familiarity a meter of ambient language? A natural experiment on effects of perceptual experience. *Memory & Cognition, 28*(5), 789–797. https://doi.org/10.3758/BF03198414

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7 http://CRAN.R-project. org/package1⁄4lme4

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. https://doi.org/10.1016/j.jml.2012.11.001

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro & O. S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Birdsong, D., Gertken, L.M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism.* COERLL, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (version 6.0.42) [computer software]. Available from www.praat.org.

Carey, M. (2004). CALL Visual feedback for pronunciation of vowels: Kay Sona-Match. *CALICO Journal, 21*(3), 571–601. https://www.jstor.org/stable/24149798

Christodoulides, G. (2016). *Effects of cognitive load on speech production and perception.*

[Unpublished doctoral dissertation]. Université Catholique de Louvain.

Chun, D. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*(1), 61–77. http://llt.msu.edu/vol2num1/article4/index.html

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

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics, 32*(4), 493–516. https://doi.org/10.1016/j.wocn.2004.05.002

Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching: a research-based approach. *TESOL Quarterly, 39,* 379–397. https://doi.org/10.2307/3588486

Garcia, C., Kolat, M. & Morgan, T. (2018). Self-correction of second language pronunciation via online, real-time, visual feedback. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference* (pp. 54–65). Ames, IA: Iowa State University.

Hammond, R. (2001). *The sounds of Spanish: Analysis and application (with special reference to American English).* Cascadilla Press.

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

Jurafsky D., Bell, A., Gregory, M., & Raymond W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure. Typological Studies in Language* (pp. 229–254). John Benjamins.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America, 138,* 817–832. https://doi.org/10.1121/1.4926561

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384–422. http://dx.doi.org/10.1080/00437956.1964.11659830

Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates training with waveform displays. *Language Learning & Technology, 13*(2), 29–47. 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. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nadeu Rota, M. (2013). *Effects of lexical stress, intonational pitch accent, and speech rate on vowel quality in Catalan and Spanish*. [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.

Offerman, H. M. (2020). *Effects of pronunciation instruction on L2 learner production and perception in Spanish: A comparative analysis.* [Unpublished doctoral dissertation]. Purdue University.

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System*, *59*, 45–60. https://doi.org/10.1016/j.system.2016.03.003

Okuno, T. (2013). *Acquisition of L2 vowel duration in Japanese by native English speakers.* [Unpublished doctoral dissertation]. Michigan State University.

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

Olson, D. J. (2019). Feature acquisition in second language phonetic development: Evidence from phonetic training. *Language Learning*, *69*(2), 366–404. https://doi-org/101111/lang.12336

Olson, D. J., & Offerman, H. M. (2020). Maximizing the effect of visual feedback for pronunciation instruction. *Journal of Second Language Pronunciation.* https://doi.org/10.1075/jslp.20005.ols

Ortega-Llebaria, M., Olson, D. J., & Tuninetti, A. (2018). Explaining cross-language asymmetries in prosodic processing: The Cue-Driven Window Length Hypothesis. *Language and Speech, 62*(4), 701–736. https://doi.org/10.1177/0023830918808823

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

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.

Saito, K. (in press). Corrective feedback and the development of L2 pronunciation. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in language learning and teaching*. Cambridge, UK: Cambridge University Press

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–159. https://doi.org/10.1093/applin/11.2.129

Thomson, R., & Derwing, T., (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326–344. https://doi.org/10.1093/applin/amu076

Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A. (2015). In U. Steindl et al., (Eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics* (pp. 64–71). Cascadilla Proceedings Project.