2019

Feature Acquisition in Second Language Phonetic Development: Evidence from Phonetic Training

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

Recommended Citation

Olson, D. J. (2019). Feature acquisition in L2 phonetic development: Evidence from phonetic training. *Language Learning, 69*(2), 366–404.

Feature Acquisition in Second Language Phonetic Development: Evidence from Phonetic Training

Daniel J. Olson

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

Abstract

This study employed a targeted phonetic instruction to explore the mechanisms that underpin second language (L2) phonetic acquisition. Broadly, two general approaches to phonetic acquisition have been previously proposed. A segmental approach suggests that learners acquire a series of individual, discrete phonemes (e.g., Flege, 1995), while a featural approach posits that L2 phonetic development occurs at the subsegmental level of the feature, which may be shared across multiple phonemes (e.g., de Jong, Hao, & Park, 2009). This study extended this line of research, using a visual feedback paradigm to train English speakers on one of the three voiceless stop consonants in Spanish. Analysis focused on the change in voice onset time across three testing sessions (pretest, posttest, delayed posttest). Results demonstrated a significant change in voice onset time for trained and nontrained phonemes, suggesting that featural changes generalize to related phonemes. Theoretical and pedagogical implications are discussed.

Keywords phonetics; second language; acquisition; feature; voice onset time; Spanish

Acknowledgements: I would like to thank Jorge Garcia, Ana Jessica Hernández Morales, Julio César López Otero, Daniela Marinho Ribeiro, and Alejandro Rodríguez de Jesús for their efforts on various parts of this project.

Introduction

As a key component in developing competency in a second language (L2), phonetics plays a large role in determining the comprehensibility, intelligibility, and accentedness of L2 speech (Munro & Derwing, 1995). In the early stages of phonetic acquisition, learners often map novel sounds in the L2 onto existing sounds in the first language (L1) (Flege, 1987). To successfully acquire some phonetic aspects of the L2, learners must establish and produce new "units" that are functionally distinct from those present in the L1. Although much work has been done on the outcomes of L2 acquisition at the phonetic level, revealing varied degrees of attainment depending on factors such as age of acquisition (e.g., Flege, 1998), intensity of exposure or input (e.g., Flege & Liu, 2001), and formal pronunciation instruction (e.g., Thompson, 1991), debate remains about the underlying mechanisms that facilitate or inhibit acquisition of the phonetic aspects of the target language.

Although a growing body of literature has begun to show a positive effect for phonetic instruction on L2 learner pronunciation (Lee, Jang, & Plonsky, 2015; Thomson & Derwing, 2015), it is unclear exactly what learners acquire. Broadly, two general approaches to phonetic acquisition can be seen in the previous literature: segmental and featural. The first approach considers that acquisition occurs on a segment-by-segment basis, in which the successful acquisition of L2 phonetic targets requires acquisition of a series of discrete, individual phonemes (e.g., Flege, 1987). The second approach posits that acquisition occurs at the level of the phonetic feature, subsegmental components that may be shared across multiple segments. Thus, acquisition of a single feature may be generalized across multiple phonemes (e.g., de Jong, Hao, & Park, 2009). Although both accounts acknowledge the role of the L1 and adequately describe outcomes in L2 speech, they differ fundamentally in the nature and size of the unit that learners acquire. From a more practical perspective, pedagogical materials (e.g., Arteaga, 2000) and classroom instruction (e.g., Foote, Trofimovich, Collins, & Soler Urzúa, 2016) often have ! not considered a featural approach and have instead focused on individual consonants and vowels (i.e., segments).

This study used a visual feedback approach to phonetic instruction (e.g., Olson, 2014b) in order to explore the mechanisms that underpin L2 phonetic acquisition. Specifically, this line of research sought to determine whether, following phonetic instruction, learners acquire (or improve production of) a single segment (i.e., a single voiceless stop) or a more generalizable phonetic feature (i.e., voice onset time (VOT)).

Literature Review

Approaches to Phonetic Acquisition

Among the most influential models in L2 phonetics, the Speech Learning Model, proposed and refined by Flege (Flege, 1987, 1988, 1991, 1995), is most concerned with the development of L2 phonetics and the role of age of exposure, as well as the nature of the interaction between learners' two phonetic systems. The Speech Learning Model posits that L2 learners are able to establish new phonetic categories regardless of age, but one of the key factors that determines whether a new category will be established is its relationship to L1 phonemes. When acquiring a new language, learners may subsume new phonemes into an already existing category (Flege, 1987), or they may create a new, separate category (e.g., Flege & Eefting, 1987). The similarity of the new L2 phoneme to existing L1 phonemes is crucial for determining the outcome. However, most relevant for the current study is the fact that such proposals are made essentially on a phoneme-by-phoneme basis. As de Jong, Hao, and Park (2009) noted, research supporting the Speech Learning Model has often examined the acquisition of phonemic contrasts in isolation. For example, Flege, Munro, and Skelton (1992) examined the production of word-final

/t/ by Mandarin-speaking learners of English, with a clear focus on the individual phonemes /t/ and /d/ rather than a broader focus on the acquisition of the voicing feature.

Although the Speech Learning Model focuses on speech production, some more perception-oriented models have mirrored this approach. Frameworks such as the Native Language Magnet theory (Kuhl 1992, 1993a, 1993b; Kuhl & Iverson, 1995) and Perceptual Assimilation Model–L2[1] (Best & Tyler, 2007) propose that perception of a new sound in the L2 is constrained by L1 phonetic categories, with the degree of similarity between the L2 and L1 category determining the perceptual outcome. Although the theoretical mechanisms differ somewhat, these models again rely more on a segmental rather than a featural level. In her summary, Brown (2000) noted that these models recognize the influence of L1 categories but generally do not provide a concrete explanation of how L2 sounds are equated with L1 sounds.

Support for a more feature-oriented approach has come most prominently from work in perception, in which the ability to perceive a feature contrast is correlated across multiple phonemes. This work, seemingly rooted in previous theoretical accounts of feature geometry (Clements, 1985; Sagey, 1986), has acknowledged that segments are comprised of several subcomponents, or features, and that several phonemes may share one or more features. As an example, the English voiceless stops /p, t, k/ all share the features of [–continuant] and [–voice] but differ in the feature of place of articulation: /p/ is [+labial], /t/ is [+coronal], and /k/ is [+dorsal]. A featural approach to acquisition postulates that learners may acquire one or more of these features rather than acquiring the segment as a whole.[2]

Within perception, the presence or absence of a featural contrast in the L1 has been used to predict perceptual abilities for a given set of phonemes in the L2 (Brown, 1997). In contrast to the Speech Learning Model, a featural account might assume that L2 perception is underpinned not by the individual segment but rather by a featural contrast that may apply across several segments. Additional support for a featural account of perception has come from monolingual ! paradigms, such as selective adaptation (Diehl, Elman, & McCuskter, 1978; Eimas & Corbit, 1973; for contrast, see Diehl, Kluender, & Parker, 1985), in which a shift in the boundary of one particular feature—VOT—at a given point of articulation was generalized across differing points of articulation. More recently, de Jong, Silbert, and Park (2009) examined correlations between discrimination accuracies across a number of consonant pairs from Korean-speaking learners of English. They found strong correlations between accuracies for consonant pairs that differed via the same featural contrast (stop–fricative). That is, participants who accurately discriminated /f/ from /p/ in the L2, a consonant pair that differs by manner of articulation, were also successful in discriminating /t/ from /θ/, which represents the same manner contrast. Moreover, de Jong et al. did not find correlations between different featural contrasts. Participants who showed high accuracy in perceiving a voicing contrast (i.e., voiced–voiceless) did not necessarily show high accuracy in perceiving a manner contrast (i.e., stop–fricative). This finding echoed the perceptual work of Brown (2000), who observed that L2 speakers were only able to perceive differences in pairs of phonemes if the contrasting feature were present in their L1. This effect persisted even when one of the phonemes was not present in the L1 inventory. Exemplifying this effect, Japanese-speaking learners of English were able to perceive the difference between /b–v/, but not /ɹ–l/. In both cases, one consonant of the pair is absent from the native inventory—/v/ and /l/, respectively, although the [+/– continuant] feature that distinguishes /b/ from /v/ is present in the learners' L1.

When examining L2 production, several authors have observed that, when a given phoneme does not exist in the L1, learners substitute a "minimally phonetically distinct" segment (Hancin-Bhatt, 1994, p. 244). In this case, minimally phonetically distinct can be defined as a segment that differs by the fewest number of features. Moreover, such features are often found in a hierarchy. This line of work was extended by de Jong, Hao, and Park (2009), who showed that ! production accuracies correlate for pairs of phonemes that share both feature contrasts and gesture contrasts. For example, performance on manner contrasts (i.e., stop–fricative) was correlated between voiced and voiceless phonemes, while performance on voicing contrasts (i.e., voiced–voiceless) was not correlated across different places of articulation. De Jong, Silbert, and Park (2009) explained these findings in terms of merging a featural approach with gestural considerations. They suggested that production requires the acquisition of discrete gestures, and such gestures are transferable across various phonemes. Successfully producing a manner contrast (stop vs. fricative) requires a similar gesture for both voiced and voiceless phonemes. In contrast, the voicing contrast (voiced vs. voiceless) requires different articulatory gestures for labial and coronal phonemes.

L2 Instruction and Instructional Research

Although the theoretical literature presents some degree of debate, current pedagogical approaches seem to implicitly take a segmental view, which is reflected in pedagogical materials and instructional practices as well as in research on L2 phonetic instruction. With respect to materials, when textbooks designed for general skills courses (i.e., not a standalone phonetics course) consider phonetics (or pronunciation), they often present information on a phoneme-by-phoneme basis. A recent review of pronunciation curricula in 48 English L2 texts found that, beyond suprasegmentals (e.g., intonation), activities commonly focused on “vowels”, “clusters”, and “consonants” (Derwing, Diepenbroek, & Foote, 2012). Further illustrating this approach, in a review of beginning-level Spanish textbooks, Arteaga (2000) noted several texts that addressed only a subset of the three voiceless stops in Spanish but not all three. This pattern, also found for voiced stops, implies that texts may consider production of bilabial and velar stops but not alveolar stops. The focus on individual segments is also reflected in L2 classroom practices (Foote et al., 2016).[3] In their survey of university-level English L2 instructors, Darcy, Ewert, and Lidster (2012) showed that, with respect to segmental production, instructors believe that focus should be on "specific consonants" and "specific vowels" (see also Breitkreutz, Derwing, & Rossiter, 2001; Foote, Holtby, & Derwing, 2011). Reference to consonants and vowels broadly suggests, by inference, a segmental approach, in that the mention of subsegmental components is largely absent. Taken as a whole, current pedagogical approaches do not tend to consider featural components, instead relying on a phoneme-by-phoneme (i.e., segmental) approach.

Furthermore, a recent review of research on L2 phonetic instruction showed a focus on training learners on individual segments rather than on broader features. In their large-scale review of research on L2 pronunciation instruction, Thomson and Derwing (2015) found that 53% of studies dealt with teaching segments and another 24% with both segmental and suprasegmental aspects. Moreover, they noted that many papers addressed single segments in isolation, such as English /ɹ/ or Spanish intervocalic /d/. This again supports the assertion that both L2 classroom practices and pedagogical research take a segmental approach as their underlying theoretical basis for L2 phonetic acquisition.

The Current Study

This study addressed the question of whether learners acquire a specific segment (i.e., phoneme) or a more generalized phonetic feature (i.e., VOT). To address this question, a phonetic training paradigm was designed and administered to several groups of language learners. Unlike acquisition via naturalistic exposure, phonetic instruction with lower-level language learners offers a unique opportunity to address this issue, as the intervention can be applied to a subset of segments that share a given feature. Two specific research questions were addressed.

1. Does training that targets one segment (i.e., phoneme) lead to significant gains for all ! segments that share a given feature (i.e., VOT)?

In this study, native English-speaking participants received training on one of the three voiceless stop consonants in Spanish (/p/, /t/, or /k/). English and Spanish differ in their realization of VOT: English has long-lag (VOT = 30–100 milliseconds) and Spanish short-lag voiceless stops (VOT = 0–30 milliseconds) (e.g., Lisker & Abramson, 1964, among many). Analysis considered the change in VOT for both the trained phoneme and the other voiceless stops not included in the training paradigm. For example, if a learner received training on /p/, did gains, defined as a reduction in VOT, generalize to the phonemes /t/ and /k/? Failure to generalize might imply a segmental interpretation of phonetic acquisition, while generalization of gains would support a featural interpretation.

2. If improvement is shown for nontarget phonemes, what is the relationship between the degree of change in VOT for trained and nontrained phonemes?

Although improvement in the production of nontrained phonemes would provide support for a featural interpretation, the relationship between improvements for trained and nontrained phonemes would speak to the strength of the link across phonemes.

Experiment 1

To answer the research questions, a targeted phonetic training paradigm using visual feedback was implemented with English-speaking learners of Spanish. Although VOT distinctions exist between English and Spanish across all three places of articulation for word-initial voiceless stops (bilabial /p/, alveolar /t/, velar /k/), participants received training on only one of the target language's three voiceless stop consonants. An oral production paradigm, in which participants produced targets with all three voiceless stop consonants embedded in utterances, was implemented prior to (pretest), immediately following (posttest), and four weeks after (delayed posttest) the phonetic training task. Analysis focused on the change in normalized VOT for both trained phonemes and their nontrained counterparts across the three sessions.

Method

Participants

Twenty-five participants were recruited from three, fourth semester Spanish courses at a large public Midwestern university. This course was an intermediate-low level course that focused on the four basic language skills (reading, writing, speaking, and comprehension) and cultural aspects of the Spanish-speaking world. All participants were placed into the fourth semester course by successfully completing the prior course in the sequence or from their score on a standardized placement exam. Following the completion of the experiment, a language background questionnaire was administered. All participants were considered to be native speakers of English and L2 learners of Spanish, having learned English from birth and Spanish after the age of 5 years (*M* = 12.90 years, *SD* = 3.76). All participants reported speaking only English in the home, growing up in English-dominant regions of the United States, and none reported any significant time in a non-English speaking region or country. The data for any participants who reported familiarity with speech analysis software or who had previously taken a course in phonetics were removed from analysis.[4]

Stimuli

Stimuli consisted of Spanish target tokens, containing word-initial voiceless stops (/p, t, k/), embedded in utterances. As was stated previously, word initial voiceless stops differ crosslinguistically in English and Spanish, with English stops being produced with long-lag VOT

(30–100 milliseconds) and Spanish being produced with short-lag VOT (0–30 milliseconds). Many authors have noted that English learners of Spanish often produce Spanish tokens with English-like VOTs (e.g., Hammond, 2001) as the result of L1 transfer. Although crosslinguistic differences in VOT may impact accentedness, they are unlikely to lead to issues of intelligibility for English learners of Spanish (Lord, 2005). Within a framework of intelligibility, comprehensibility, and accentedness, many authors (e.g., Munro, 2016, among others) have argued for a pedagogical focus on features that impact intelligibility. However, VOT was chosen for both its theoretical value and pedagogical implications. Namely, VOT provides a gradient (versus categorical) measure and has been shown to improve following visual feedback (Offerman & Olson, 2016). Moreover, VOT has been shown to impact intelligibility in other language pairings, incuding L1 Spanish–L2 English (Hualde, 2005).

A total of 90 two-syllable target words, 30 for each voiceless stop, were included in the experimental design. Given the role that cognate status may play in VOT (Amengual, 2012), all tokens were noncognate. In addition, given that stress impacts VOT production (see Lisker & Abramson, 1967), all targets were controlled for stress placement, with stress on the initial syllable (i.e., paroxytonic). As vowel height has been shown to impact VOT (e.g., Flege, 1991), target words were balanced for the vowel following the initial stop, such that each initial stop was followed by either the mid vowel /o/ or low vowel /a/. Each resulting consonant–vowel (CV) combination was represented by 15 unique words (3 initial stops × 2 vowels × 15 words = 90 target tokens). Stimuli were divided across three sessions (pretest, posttest, and delayed posttest), with 10 instances of each initial stop in each session. All target words were placed in utterance medial position, and there were no occurrences of any of the voiceless stops in the portion of the utterance preceding the target token. Each voiceless stop was immediately preceded by a mid vowel (/e/ or /o/). Table 1 provides sample utterances containing the target tokens. An additional

30 filler tokens were included. All filler tokens began with the orthographic <h> (see Experiment ! 2 for discussion).

Table 1 Sample stimuli for voiceless stops

Initial segments	Sample stimuli !
/pa/	*Miles de patos nadan en los lagos grandes.* & "Thousands of ducks swim in the big lakes." !
/ta/	*Si llego tarde mañana, mi madre me gritará.* & "If I arrive late tomorrow, my mother will yell at me." !
/ka/	*Mi abuelo vende camas en su tienda.* & "My grandfather sells beds in his store." !

Note. Target tokens are underlined.

As the training paradigm focused on words in isolation, a unique set of stimuli, consisting of 40 words in isolation, balanced for initial consonants (/p, t, k/ and *h*) and following vowels /i, e/ were also recorded. Unlike the targets embedded in utterances, the same words in isolation were repeated in each of the three recording sessions. Although these tokens served to validate the effect of training, such analysis was considered secondary given that the tokens in connected utterances represented a more natural task (see Thomson & Derwing, 2015).

Familiarity and Frequency Norming Study

Given the strict constraints used to choose stimuli, the target words varied in their respective frequencies. As relative frequency has been shown to impact phonetic production (e.g., Jurafsky, Bell, Gregory, & Raymond, 2001), it was important to ensure a balanced distribution of target token frequencies across the three experimental sessions. Because standard frequency measurements are based on native speaker corpora, they may be less suitable for L2 learners. Consequently, a separate subjective familiarity rating procedure was conducted. This familiarity norming study allowed for an equal distribution of token familiarity across the three experimental sessions.

Participants for the norming study (*N* = 19), different from those participating in the larger study, were drawn from a similar population. All participants were students in a fourth semester Spanish class. They were native English speakers and reported speaking exclusively English in the home and learning Spanish after the age of 5 years (*M* = 13.70 years, *SD* = 1.32). Participants were presented with a randomized list of target tokens and fillers and asked to rate each one based on their own personal familiarity with the word. Verbs were always presented in the infinitive form, and nouns and adjectives in the singular form. Participants were asked to rate their familiarity with the items on a 7-point Likert scale with fully labeled intervals (Auer, Bernstein, & Tucker, 2000; Nusbaum, Pisoni, & Davis, 1984). Table 2 provides the scale labels used for familiarity ratings.

Table 2 Word familiarity rating scale for norming study for Experiment 1

Scale	Label
1	I have never seen or heard this word and I don't know its meaning. !
2	I might have seen or heard this word, but I don't know its meaning. !
3	I'm pretty sure I have seen or heard this word, but I don't know its meaning. !
4	I have seen or heard this word before, but I don't know its meaning. !
5	I am sure I have seen or heard this word before, but I have only a vague idea ! of its meaning. !
6	I am sure I have seen or heard this word before, and I think I know the ! meaning, but I'm not sure it's correct. !
7	I know the word and am confident of its meaning. !

Average familiarity ratings were calculated for each individual token. Across all target tokens, participants reported a moderate familiarity with the words (*M* = 4.50, *SD* = 1.85). Individual tokens ranged from highly unfamiliar (e.g., *pomos* "door knobs," *M* = 1.53, *SD* = 1.02) to the highly familiar (e.g., *casa* "house," *M* = 7.00, *SD* = 0.00). To assess the familiarity of the stimuli used in each of the three sessions, ANOVAs were conducted on the familiarity ratings for target tokens included in each of the three different recording sessions. Separate ANOVAs were conducted for the different word-initial phonemes. Results demonstrated no significant difference in word familiarity across each of the three sessions for any of the three phonemes: /p/ $F(2, 27) = 0.029$, $p = .972$, $\eta^2 = .002$; /t/ $F(2, 27) = 0.004$, $p = .996$, $\eta^2 < .001$; /k/ $F(2, 27) = 0.001$, $p = .999$, $\eta^2 < .001$. In short, although tokens presented a range of familiarity, there was no difference in familiarity within each phoneme between any of the three sessions.

Procedure

The phonetic training task used in this study was a visual feedback paradigm (Olson, 2014a). Visual feedback consists of presenting learners with a visual representation of their productions and allowing them to compare their productions to native speaker productions. Early visual feedback paradigms presented learners with intonation contours for training on suprasegmental features (e.g., de Bot, 1980), and subsequent iterations of the paradigm have presented learners with spectrograms and/or waveforms (e.g., Saito, 2007) and schematic representations of acoustic differences (e.g., Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015) for training them on segmental features.[5] Visual feedback has been shown to be successful for teaching a variety of segmental features, including vowel length (Okuno, 2013), singleton–geminate contrast (Motohashi-Saigo & Hardison, 2009), vowel formant accuracy (Saito, 2007), and intervocalic consonantal lenition (Olson, 2014a). Visual presentation of waveforms and spectrograms was chosen for this study because it has been successfully used to improve VOT (Offerman & Olson, 2016). The visual feedback paradigm consisted of several phases: (a) prerecording, (b) self and native speaker analysis, (c) nonnative and native speaker comparison, and (d) rerecording. The analyses and comparisons were conducted during the course of one 50-minute class meeting.

For the prerecording phase (i.e., pretest), participants recorded the given set of stimuli, including both words in isolation and the targets and fillers embedded in utterances. First, participants sent digital copies of the recordings to the course instructor. Second, using Praat

(Boersma & Weenink, 2017), participants were instructed to print out the "visual representation" ! (i.e., waveform and spectrogram) of the first five words produced in isolation for in-class analysis. These first five words all began with the target phoneme for that particular experimental group. They also were asked to attempt to segment the word into individual "sounds" (i.e., phonemes), mainly through repeated listening. This prerecording phase was conducted at home on students' personal computers, was included in the curriculum as required class homework, and was graded on a complete or incomplete basis. Recording instructions were given in the target language.

During the self-analysis phase, participants answered a series of questions about the images of their own productions, with a focus on the target segment. Example 1 shows the guiding questions for the group who received training on the voiceless bilabial stop /p/, translated from the L2. During the native speaker analysis phase, participants were given a spectrogram and waveform produced by a native speaker (male, peninsular variety) and prompted with a parallel set of questions. Figure 1 shows the waveform and spectrogram of the word *peña* "group" or "club" produced by a L2 speaker (left panel) and a native speaker of Spanish (right panel).

Example 1

a. How did you decide where the boundaries of each sound were?

b. What are the visual characteristics of your *p*?

c. Is your *p* longer or shorter than the *e*?

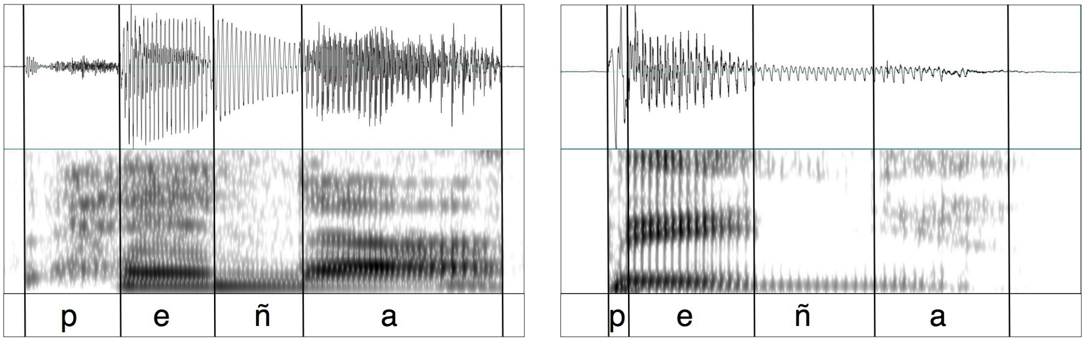

Figure 1 Comparison of waveforms and spectrograms of *peña* "group/club" produced by a L2 speaker (left) and a native speaker of Spanish (right).

To promote native and nonnative speaker comparison, participants were given a set of guiding questions asking them to describe the visual difference between native and nonnative speaker productions and to hypothesize about the auditory differences. They were then given auditory recordings of a native and nonnative speaker to confirm their hypotheses. Example 2 gives guiding questions for both visual and auditory comparison. As further practice, participants were given several pairs of spectrograms for novel words and asked to identify which word in the pairs was produced by a native Spanish speaker.

Example 2

a. Describe the visual difference between your *p* and the *p* produced by a native Spanish speaker.
b. What do you think the auditory difference is between your *p* and the *p* produced by a native Spanish speaker?
c. Listen to the following pair of words, the first was produced by a nonnative speaker and the second by a native Spanish speaker. How would you describe the auditory difference?

Following the intervention, participants were given three days to record the second set of stimuli at home (i.e., posttest). These stimuli included the same set of words in isolation and a unique set of words in utterances. The delayed posttest, again including the same words in isolation and a third unique set of target words in utterances, was conducted approximately four weeks after the intervention (for delayed posttest timing in L2 acquisition research, see Norris & Ortega, 2000). The delayed posttest occurred during the same semester, which limited the likelihood that participants experienced a drastic shift in their usage or exposure patterns (e.g., study abroad). Moreover, the coursework during the intervening period was similar for all groups. Although such exposure data were lacking, it was anticipated that all groups received similar amounts of exposure to the target language between the posttest and delayed posttest. For both the posttest and delayed posttest, instructions, procedures, and grading rubric (i.e., complete or incomplete) paralleled those employed in the pretest.

To allow for a balanced experimental approach, each of the three groups (i.e., classes) received training on one of the three Spanish stop consonants: /p/, /t/, or /k/. Of the 25 participants, eight received training on /p/, 13 received training on /t/, and four received training on /k/. Although each group received training on one stop consonant, they recorded stimuli that contained all three of the relevant phonemes. Unequal group sizes resulted from differences in class size and from the number of participants who failed to meet the inclusionary criteria. Mitigating this difference, groups were collapsed during analysis. The visual feedback paradigm was included in the course curriculum and students received a grade based on completion of all parts of the training, although providing data for the current project was voluntary. Participants received no feedback on their pronunciation and received no compensation for their participation.

Data Analysis

A total of 2,250 tokens were included in the initial analysis (25 participants × 3 initial phonemes ! × 10 tokens × 3 sessions = 2,250 tokens). Of those, approximately 4.6% of tokens (*k* = 104) were eliminated due to various production and recording errors, including missing data, yawning, laughter, and poor recording quality. Additionally, outliers two standard deviations above and below the mean (5.42%, *k* = 122) were eliminated, with a total of 2,024 tokens retained for the final analysis.

VOT was measured using Praat (Boersma & Weenink, 2017) and defined as the temporal difference between the release of the oral closure (i.e., burst) and the onset of vocal fold vibration (i.e., periodic waves). Crosslinguistically, VOT varies across different places of articulation, with bilabials evidencing the shortest VOTs and velars the longest, although the reasons for such variation are subject to debate (for review, see Cho & Ladefoged, 1999). For this study, VOT values were normalized to allow for direct comparison across places of articulation. To normalize the values, the VOT for each token was converted into a ratio based on the average Spanish and English VOT values from the seminal work by Lisker and Abramson (1964). In this ratio, a value of 0 represents a token with a VOT equal to that of the average Spanish VOT (/p/ = 4 milliseconds; /t/ = 9 milliseconds; /k/ = 29 milliseconds). A value of 1 represents a token with VOT equal to that of the average English VOT (/p/ = 58 milliseconds; /t/ = 70 milliseconds; /k/ = 80 milliseconds). Normalizing VOT values also allowed for the different groups to be collapsed for analysis.

Initial statistical analysis consisted of a linear mixed-effects model examining normalized VOT values, conducted using R statistical software (R Core Team, 2013) and the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Fixed effects included both session (pretest, posttest, delayed posttest) and token type (trained, nontrained). Subject was included as a random effect, with both random slopes and intercepts for each of the main factors and their interactions (see

Barr, Levy, Scheepers, & Tily, 2013). The significance criterion was set at |*t*| = 2.00. Standard effect sizes (Cohen's *d*), calculated independently from the mixed-effects models, included pooled standard deviations. Confidence intervals for the effect sizes were calculated using the psych package (Revelle, 2018) for R.

Results

Words in Utterances

Results for the initial model (Table 3) demonstrated a significant effect of session on the production of normalized VOT. Specifically, although the improvement between the intercept (pretest, trained) (*M* = 0.69 , *SD* = 0.41) and the posttest (*M* = 0.61 , *SD* = 0.38) was not significant, the difference between the pretest and delayed posttest (*M* = 0.59, *SD* = 0.35) was significant. There was no difference between the posttest and delayed posttest (not listed in Table 3), *b* = –0.035, *t* = –0.800, *d* = 0.08, 95% CI [–0.18, 0.34]. For token type, there was no significant difference between the trained (*M* = 0.69, *SD* = 0.41) and nontrained (*M* = 0.79, *SD* = 0.40) phonemes at the pretest, implying that both types of phonemes were produced similarly before training. However, there was also no significant interaction between session and token type. In sum, as Figure 2 shows, although VOT decreased following instruction, the change in VOT was similar for trained and nontrained phonemes.

Table 3 Linear mixed-effects model results for comparison of trained and nontrained phonemes in Experiment 1

Parameters	*b*	*SE*	95% CI	*t*	*d*	*95% CI*
Intercept (pretest, trained)	0.704	0.068	[0.568, 0.840]	10.288		
Posttest	–0.084	0.047	[–0.178, 0.010]	–1.799	0.17	[–0.09, 0.42]
Delayed posttest	–0.119	0.059	[–0.237, –0.001]	–2.005	0.28	[0.02, 0.53]
Non-trained	–0.011	0.031	[–0.073, 0.051]	–0.370	0.03	[–0.28, 0.23]
Posttest: Nontrained	0.002	0.037	[–0.072, 0.076]	0.041	0.18	[–0.08, 0.43]
Delayed posttest: Nontrained	0.011	0.046	[–0.081, 0.103]	0.241	0.25	[–0.00, 0.51]

To further support collapsing the different groups into one group, an additional mixed-effects model was conducted, parallel to the first, but it included group as a fixed effect. The group factor was defined by the phoneme on which each subject received training, /p/, /t/, or /k/. The resulting model was then compared to the initial model. Relative to the main model (Akaike Information Criterion = 640.04), the inclusion of group as a fixed effect (Akaike Information Criterion = 653.25) did not lead to any significant improvement in model fit, $\chi^2(12) = .547$. Thus, participants performance was similar regardless of whether they had received training on /p/, /t/, or /k/,

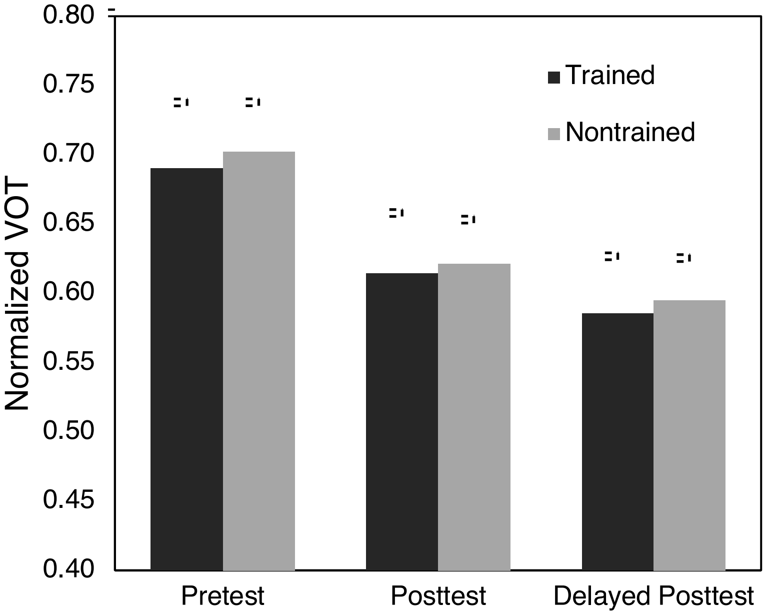

Figure 2 Normalized voice onset time (VOT) ratios for trained and nontrained phonemes across all three sessions in Experiment 1. Error bars represent +/–1 SE.

To contextualize these findings, as VOT is often presented in milliseconds, it is worth considering raw, nonnormalized data. As an example (Figure 3), the group trained on the phoneme /t/ evidenced an average decrease of 7.29 milliseconds in the VOT of /t/ from the pretest (*M* = 54.18 milliseconds, *SD* = 23.83) to the delayed posttest (*M* = 46.89 milliseconds, *SD* & = 20.98). This group also demonstrated a change in VOT for the two voiceless stops on which it had received no training: an 8.62 milliseconds decrease for /p/ from the pretest (*M* = 47.58 milliseconds, *SD* = 21.94) to the delayed posttest (*M* = 38.95 milliseconds, *SD* = 17.15), and an 7.97 millisecond decrease in the VOT of /k/ from the pretest (*M* = 65.66 milliseconds, *SD* = 19.58) to the delayed posttest (*M* = 57.69 milliseconds, *SD* = 14.85). Although the final mean VOT values produced for all phonemes (/t/= 46.89 milliseconds, /p/ = 38.95 milliseconds, /k/ = 57.69 milliseconds) did not fall within the accepted range for Spanish VOT, they had shifted towards a more Spanish-like production.[6] Nonnormalized raw results for the other two instructional groups are included in Appendix S1 in the Supporting Information online.

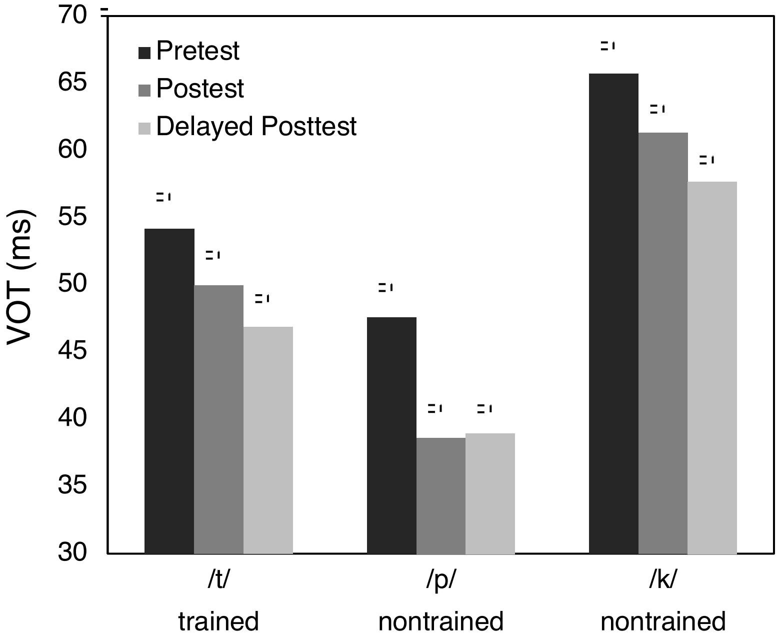

Figure 3 Raw voice onset time (VOT) values in milliseconds for each voiceless stop produced by the group receiving training on /t/in Experiment 1. Error bars represent +/–1 SE.

To address the relationship between the degree of change in VOT for trained and nontrained phonemes, a subsequent statistical analysis was performed to investigate the relationship between the change in VOT for the trained phonemes relative to the nontrained phonemes. ΔVOT (change) values were computed for each participant's trained and nontrained phonemes by subtracting the average normalized VOT value at the delayed posttest from the normalized VOT value at the pretest. A linear regression was then conducted, using the R lm function (R Core Team, 2013), to compare the ΔVOT values for trained and nontrained phonemes. Results demonstrated a strong relationship between the change in VOT for trained and nontrained phonemes, $R^2 = .42$, $F(1, 23) = 16.63$, $p < .001$, $b = 0.941$. The slope of the regression line approximated 1.00, demonstrating that a change in normalized VOT for trained phonemes was matched by a similar change in nontrained phonemes (Figure 4).

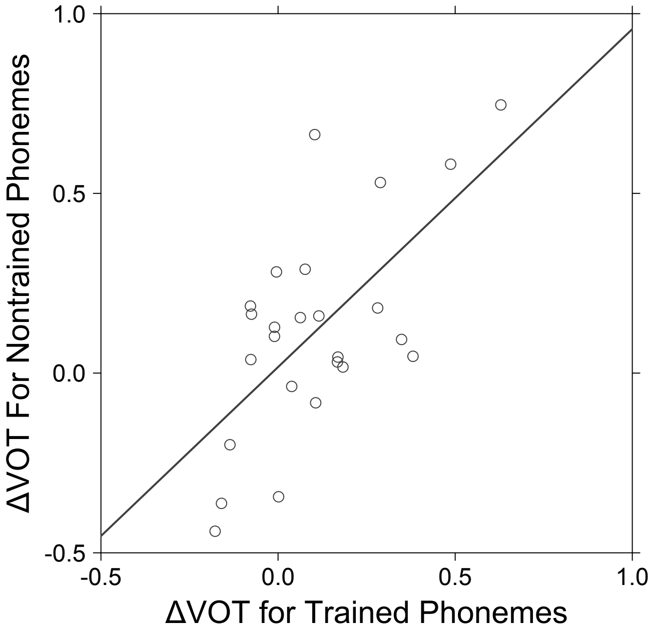

Figure 4 Scatterplot for the change in normalized voice onset time (VOT) values for trained versus nontrained phonemes for participants in Experiment 1, with a linear regression line showing the best fit to the data.

Words in Isolation &

Although the primary analysis focused on the novel tokens produced within connected speech, it was worth considering the results for the words produced in isolation as a complementary analysis. Although farther removed from natural connected speech and the related cognitive load because the visual feedback paradigm focused on tokens in isolation, an examination of such tokens served to validate the impact of the training paradigm. After eliminating errors (*k* = 91) and outliers (*k* = 86), a total of 1,982 tokens were included in the analysis.[7] The statistical analysis employed mirrored the initial analysis for tokens in connected speech. The results found for the words produced in isolation closely paralleled the results found for the tokens produced within connected speech. There was a significant effect of session, with differences shown between the intercept, that is, pretest, nontrained (*M* = 0.68, *SD* = 0.36), and both the posttest (*M* = 0.59, *SD* = 0.33), *b* = –0.102, *t* = -2.391, *d* = 0.26, 95% CI [0.08, 0.45], and delayed posttest (*M* = 0.59 milliseconds, *SD* = 0.35), *b* = -0.101, *t* = –2.808, *d* = 0.28, 95% CI [0.09, 0.46]. Moreover, there was no significant interaction between session and type (|*t*| < 2.00), implying that the training paradigm impacted similarly both trained and nontrained phonemes. Full descriptive statistics for words in insolation are included in Table 4. These findings further linked the improvement found in connected speech to the visual feedback paradigm.

Table 4 Normalized mean and standard deviation values (in parentheses) for voice onset time in words spoken in isolation in Experiment 1

Test	Trained phonemes	Nontrained phonemes
Pretest	0.67 (0.38)	0.68 (0.36)
Posttest	0.60 (0.33)	0.59 (0.33)
Delayed posttest	0.63 (0.34)	0.59 (0.35)

Summary of Results

Results demonstrated a significant impact of session on the production of VOT, such that VOT ! was significantly reduced from the pretest through the delayed posttest. Although effect sizes !

were relatively small (*d* = 0.28, 95% CI [0.02, 0.53]), this was not unexpected given the short ! duration of the training. As a comparison, Offerman and Olson (2016) found reductions in VOT of approximately 20 milliseconds following a series of three visual feedback paradigm trainings (effect size not available), whereas the current study showed reductions of approximately 8 milliseconds. However, most important for the current study, there was no significant difference between the amount of change in VOT for the trained and nontrained phonemes. This finding was reinforced by similar effect sizes for pretest and delayed posttest comparisons for trained (*d* = 0.28, 95% CI [0.02, 0.53]) and nontrained phonemes (*d* = 0.29. 95% CI [0.10, 0.47]). Furthermore, results from a linear regression demonstrated a strong link between the degree to which individual participants experienced a change in their trained and nontrained phonemes.

Experiment 2

In light of the results of Experiment 1, showing that participants improved on both trained and nontrained phonemes, it seemed possible to attribute this improvement to a more generalized improvement in production not specifically limited to VOT or resulting from the visual feedback paradigm. In other words, pronunciation may have improved over time, or the inclusion of a training component may have led participants to focus on other features beyond the phoneme in question. To investigate this possibility, as second experiment was conducted. The second experiment was parallel to Experiment 1, with one key exception. Participants in Experiment 2 received training on the pronunciation of *h*, the initial grapheme in each of the filler words in Experiment 1. Analysis again focused on the production of VOT, and improvement on the voiceless stops would imply a general pronunciation improvement across the three recording sessions.

Method *

Participants

Seven participants, different from those who participated in Experiment 1, were included in Experiment 2. All participants were students in a fourth semester Spanish course. Participants were native speakers of English, spoke only English in the home, and had not spent any significant time in a non-English speaking country. All participants had learned Spanish after the age of 5 years (*M* = 16.80 years, *SD* = 2.92). None had taken a course in phonetics or used speech analysis software previously.

Stimuli

Stimuli for Experiment 2 were the same as those for Experiment 1. The 30 filler tokens for Experiment 1, all words beginning with an orthographic *h*, served as the basis for training in Experiment 2. Although <h> in English usually corresponds to the voiceless fricative /h/, <h> in Spanish is never pronounced. This crosslinguistic mismatch often leads English-speaking L2 learners of Spanish to "mispronounce" this segment (e.g., Morgan, 2010). Unlike other crosslinguistic differences, for example, those involving vowels (for discussion, see Olson, 2014a), the difference between English and Spanish with respect to <h> is likely to be visually intuitive when spectrograms and waveforms are examined, given that <h> corresponds to a period of frication in English and has no articulation (frication or otherwise) in Spanish. Given both the likelihood of mispronunciation and the visual distinction between English and Spanish, <h> was chosen as a reasonable target for comparison. All of the <h> initial words followed the same constraints as the voiceless stops: two-syllable, paroxytonic, and noncognate words with orthographic <h> in initial position. With respect to phonetic context, <h> was followed by /o/ or /a/ and preceded by the mid vowels /e/ or /o/. All tokens were subjected to the same word familiarity norming procedure, and results demonstrated no significant difference in familiarity between the three sessions, $F(2, 27) = 0.045$, $p = .989$, $\eta^2 < .001$.

Procedure

The training procedure used in Experiment 2 was identical to that of Experiment 1, with one exception. Instead of participants receiving training via visual feedback on one of the voiceless stops, they were given a visual feedback paradigm addressing the correct pronunciation of *h*.

Data Analysis

For comparison to Experiment 1, data analysis focused on the normalized production of the three voiceless consonants /p, t, k/. A total of 630 tokens were included in the analysis (7 participants × 3 initial phonemes × 10 tokens × 3 sessions = 630 tokens). Approximately 5.9% of the data were eliminated because of production errors (14) and outliers (23), for a total of 593 tokens included in the statistical analysis.

Results

Words in Utterances

Paralleling the analysis employed in Experiment 1, statistical analysis consisted of a linear mixed-effects model on the dependent variable of normalized VOT values, with session as a fixed effect and subject as a random effect with random slope and intercept. Results for the initial model, shown in Table 5, stand in contrast to those found in Experiment 1. Specifically, there was no significant difference between the pretest ($M = 1.03$, $SD = 0.46$) and either the posttest ($M = 1.02$, $SD = 0.43$), $b = –0.013$, $t = –0.192$, $d = 0.02$, 95% CI [–0.25, 0.30], or delayed posttest ($M = 0.97$, $SD = 0.44$), $b = –0.047$, $t = –0.923$, $d = 0.13$, 95% CI [–0.15, 0.40].

That is, participants in Experiment 2 who had received training on <h> showed no significant ! change in their production of voiceless stops. Given the smaller relative size of the group in Experiment 2 and the above nonsignificant results, a power analysis was conducted using a simulation-based approach in the simr package (Green & MacLeod, 2016) for R. Results showed that the current model design, with an estimated size of *b* = –.16 for the fixed effect session (based on the matched experimental group), surpassed the 80% power threshold. This finding suggests that the nonsignificant findings were not the result of an underpowered design.

Table 5 Initial linear mixed-effects model for the control group who received training on <h> in Experiment 2

Parameters	*b*	*SE*	95% CI	*t*	*d*	*95% CI*
Intercept (control, pretest)	1.024	0.141	[0.742, 1.306]	7.265		
Posttest	–0.013	0.069	[–0.151, 0.125]	–0.192	0.02	[–0.25, 0.30]
Delayed posttest	–0.047	0.051	[–0.149, 0.055]	–0.923	0.13	[–0.15, 0.40]

A second model was conducted to compare the normalized VOT of the <h> group (i.e., control group) to the results for the nontrained phonemes for participants in Experiment 1 (i.e., experimental group). For this model, session and group were included as fixed effects and subject as a random effect with random slopes (by group) and intercepts. The random effects structure was simplified relative to that of Experiment 1 to permit model convergence. Full results are available in Appendix S2 in the Supporting Information online. Although the interaction between group and session was significant at the posttest, *b* = –0.077, *t* = –2.115 and nearly significant at the delayed posttest, *b* = –0.068, *t* = -1.866, there was a significant difference between the two groups at the pretest, *b* = –0.326, *t* = –2.482. Specifically, the control group (*M* = 1.02, *SD* = 0.46) produced significantly longer VOTs than the experimental group (*M* = 0.69, *SD* = 0.41), suggesting that the two groups may not have been well-matched prior to the training procedure.[8]

To better match the two groups, an experimental subgroup was identified by splitting the participants from Experiment 1 into two separate groups based on pretest performance for nontrained phonemes. The experimental subgroup (*n* = 12) consisted of those participants who produced longer mean normalized VOTs than the group median (*Mdn* = 0.765) at the pretest. A parallel analysis compared the experimental subgroup to those in the control group (Table 6). The model parameters were identical to those of the initial model of Experiment 2. The analysis demonstrated that the control group (*M* = 1.03, *SD* = 0.46) and the experimental subgroup (*M* = 0.97, *SD* = 0.30) were not significantly different at the pretest, suggesting that the control group and experimental subgroup were a better match. Results for the control group demonstrated no significant difference between the intercept (control group, pretest) and either the posttest (*M* = 1.02, *SD* = 0.43), or delayed posttest (*M* = 0.97 milliseconds, *SD* = 0.44). This result was expected given the initial analysis for Experiment 2. In contrast, the experimental subgroup results (not shown in Table 6), demonstrated a significant difference between the intercept (*M* = 0.97, *SD* = 0.30) and the posttest (*M* = 0.79, *SD* = 0.33, and the delayed posttest (*M* = 0.76, *SD* = 0.32).

Table 6 Linear mixed-effects model results for comparison of experimental subgroup and control group in Experiment 2

Parameters	*b*	*SE*	95% CI	*t*	*d*	*95%*
Intercept (control, pretest)	1.023	0.120	[0.783, 1.263]	8.504 !		
Posttest	–0.015	0.032	[–0.079, 0.049]	–0.467	0.02	[–0.2
Delayed posttest	–0.047	0.032	[–0.111, 0.017]	–1.496	0.13	[–0.1
Experimental subgroup	–0.051	0.130	[–0.311, 0.209]	–0.394	0.15	[–0.1
Posttest: Experimental subgroup	–0.169	0.045	[–0.259, –0.079]	–3.787	0.60	[0.3
Delayed posttest: Experimental subgroup	–0.160	0.045	[–0.25, –0.07]	–3.567	0.66 !	[0.3

Of most interest for this study was a significant interaction between session and group at both the posttest, *b* = –0.169, *t* = –3.787, and the delayed posttest, *b* = –0.160, *t* = –3.567. These results demonstrated that the Experiment 1 subgroup significantly reduced their VOTs, but the group in Experiment 2 did not. Moreover, although this effect held for the full experimental group, it was even more pronounced for the experimental subgroup that was more closely matched for initial performance (Figure 5).

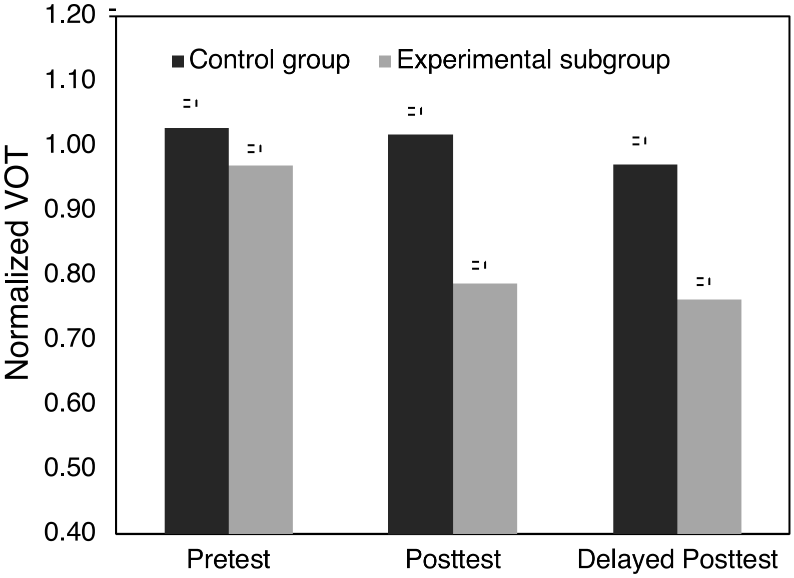

Figure 5 Normalized voice onset time (VOT) ratios for voiceless stops produced by the control and matched experimental subgroups across all three sessions in Experiment 2. Error bars represent +/–1 SE.

Words in Isolation

An analysis of the words in isolation is included as complementary to the main analysis of words produced in connected speech. A total of 483 tokens were included in the analysis of words in isolation for the control group. As in Experiment 1, results for words in isolation largely paralleled those found for connected speech. As they had with the words embedded in utterances, the participants in the control group (*M* = 0.81, *SD* = 0.37) produced longer, albeit not significantly, normalized VOTs at the pretest than the experimental group (*M* = 0.68, *SD* = 0.36). In an analysis parallel to the one above, a comparison between the experimental subgroup and control group using a mixed-effects model showed a significant difference between the intercept for the control group at the pretest (*M* = 0.81, *SD* = 0.37) and the posttest (*M* = 0.72, *SD* = 0.35), *b* = –.093, *t* = –3.124, *d* = 0.23, 95% CI [–0.08, 0.55], and the delayed posttest (*M* = 0.75, *SD* = 0.35), *b* = –0.078, *t* = –2.623, *d* = 0.15, 95% CI [–0.16, 0.46]. Moreover, there was a clear trend towards an interaction between session and group: for the experimental group at the posttest (*M*

= 0.70, *SD* = 0.31), *b* = –0.067, *t* = –1.843, *d* = 0.30, 95% CI [–0.01, 0.62], and the experimental ! group at the delayed posttest (*M* = 0.73, *SD* = 0.32), *b* = –0.063, *t* = –1.746, *d* = 0.22, 95% CI [–0.09, 0.54],.[9] Descriptive statistics for words in insolation for both the control group and the experimental subgroup are provided in Table 7. Taken as a whole, the pattern of results was similar for both the tokens in utterances and in isolation. Given that the training paradigm was implemented using words in isolation, this analysis links the reduction in normalized VOT for tokens in utterances with the training paradigm.

Table 7 Normalized mean and standard deviation values (in parentheses) for voice onset time in nontrained words spoken in isolation in Isolation in Experiment 2

	Control group	Experimental subgroup
Pretest	0.81 (0.37)	0.88 (0.28)
Posttest	0.72 (0.35)	0.70 (0.31)
Delayed posttest	0.75 (0.35)	0.73 (0.32)

Summary of Results

The results from Experiment 2, in which participants received training on the commonly mispronounced feature of *h*, stand in contrast to those found in Experiment 1. Although participants in Experiment 1 showed significant improvements for both trained and nontrained voiceless phonemes, those in Experiment 2 demonstrated no significant shift in their productions of VOT (at delayed posttest, *t* = -1.496, *d* = 0.13, 95% CI [–0.15, 0.40]). In addition, a more carefully matched subgroup of participants from Experiment 1 demonstrated relatively larger effects of the training paradigm on the nontrained phonemes (at delayed posttest, *t* = -6.522, *d* = 0.66, 95% CI [0.37, 0.94]). Although effect sizes are informative, caution should be exercised with comparisons to predetermined effect size benchmarks (e.g., Cohen, 1988; Plonsky & Oswald, 2014). A number of factors may have mitigated the size of the effects in this study, including the short length of treatment (see Offerman & Olson, 2016), the in-class (rather than laboratory) paradigm, and the focus on consonantal segments (for discussion of effect sizes in pronunciation instruction research, see Lee et al., 2015). Moreover, it should be noted that the main goal of this study was not necessarily to maximize the benefit of phonetic training but rather to examine the potential for links between phonemes at the abstract feature level. Thus, even relatively small effects, particularly for nontrained phonemes, may be informative. Taken together, results from Experiment 2 suggest that the improvements for nontrained phonemes in Experiment 1 are unlikely to be the result of a more holistic improvement in pronunciation, derived either from the phonetic training task or ongoing exposure to the target language. Instead, changes in the nontrained phonemes in Experiment 1 are likely intrinsically linked to the trained phonemes.

Discussion

Support for Feature Acquisition in L2 Phonetics

This study adds to the ongoing debate regarding the underlying nature of L2 phonetic acquisition. As native English-speaking participants selectively received training (via a visual feedback paradigm) on one of the three voiceless stops in Spanish, analysis focused on both trained and related nontrained phonemes. Related directly to the first research question, which asked whether training that targets one segment leads to significant gains for all segments that share a given feature, the results from Experiment 1indicated that training one single voiceless stop led to significant improvement across the whole set of voiceless stops. This generalization was found regardless of the place of articulation of the targeted phoneme. Moreover, responding to the second research question, which targeted the relationship between the degree of change for trained and nontrained phonemes, the degree of improvement for trained and nontrained phonemes was found to be strongly, positively correlated. In Experiment 2, in which training focused on the unrelated, commonly mispronounced grapheme *h*, participants showed no improvement in VOT. Thus, the change in VOT production for nontrained phonemes found in Experiment 1 can be reliably attributed to the effects of the visual feedback paradigm rather than to more holistic improvements in L2 phonetic production.

Within theoretical approaches to the mechanisms underlying L2 phonetic acquisition, two general lines have been previously identified—segmental and featural. In the segmental approach, which has been tacitly adopted by a number of models, including the Speech Learning Model (Flege, 1987, 1988, 1995), Native Language Magnet theory (Kuhl, 1992), and Perceptual Assimilation Model–L2 (Best & Tyler, 2007), acquisition of L2 phonetic norms occurs on a segment-by-segment basis. This approach is reflected both in the broader theoretical description, in which the relationship of the L2 sound to the L1 sound is among the key factors that determine the success with which a given sound will be acquired, as well as the methodological approach in which studies had focused on the acquisition of a single phoneme rather than a class of phonemes (for discussion, see de Jong, Silbert, & Park, 2009). In contrast, within a featural approach to L2 phonetic acquisition, rooted in feature geometry (Clements, 1985; Sagey, 1986), learners may acquire a feature or set of features that apply to multiple phonemes. Not only is the relationship between a given sound in the L2 and L1 important in a featural approach, the relationship between sounds within a given language is also crucial. The results from this study seem to support a more feature-oriented approach in that a shift in the VOT for one phoneme was generalized to other phonemes that share the same feature. In addition, the degree of change in the nontrained phonemes was strongly correlated with the change in the trained phonemes. If acquisition of this feature were limited or compartmentalized to a single phoneme, one might expect minimal shift for the nontrained phonemes. The results suggest little phoneme specificity, with the featural change generalizing in a symmetrical manner across all related phonemes.

This feature-oriented approach has found recent support in the L2 literature from both perception and production paradigms. For example, de Jong, Silbert, and Park (2009) found correlations between perceptual accuracies for multiple phoneme pairs that differed by the same featural contrast (i.e., for stop–fricative contrast, accuracies in discriminating /p–f/ correlated strongly with accuracies for /t–θ/). However, there was no correlation found between discrimination accuracies for phoneme pairs that employed different featural contrasts, such as stop–fricative versus voiced–voiceless. These results implied that learners may acquire a featural contrast and apply that "skill" to all related phonemes. Likewise, de Jong, Hao, and Park (2009) broadly found correlations in production performance across phoneme pairs that differed by the same feature, albeit with some consideration for gestural differences.

Other studies have shown that processes that impact VOT for a single stop consonant generalize to other voiceless stop consonants. One such case is selective adaptation, in which the perception of stimuli along a contrast continuum can be shifted following repeated exposure to one of the continuum endpoints. In their seminal study, Eimas and Corbit (1973) demonstrated that, when performing a perceptual categorization of synthetic stimuli from a /pa/–/ba/ continuum, participants were more likely to perceive stimuli as /pa/ following repeated presentation of /ba/ syllables (see also Samuel, 1986; Vroomen, van Linden, Keetels, de Gelder, & Bertelson, 2004; Vroomen, van Linden, de Gelder, & Bertelson, 2007); for these participants, the voiced–voiceless boundary had shifted towards the adapting stimulus (i.e., /ba/ in this example). Relevant for the current study, this process of adaptation generalized to other places of articulation that had not been included in the adaptation process, such that repeated presentations of the /ba/ stimulus also shifted the boundary in a /ta/–/da/ continuum. Suggesting a more universal role for features, Kuhl and Miller (1978) found that even some animals (e.g., chinchillas) are capable of generalizing the effects of selective adaptation across differing points ! of articulation.

A complementary process, known as recalibration or retuning, is the process by which listeners rapidly adjust their phonetic categories in response to novel accents or realizations of a given phoneme (see Norris, McQueen, & Cutler, 2003). For example, when speakers are exposed to ambiguous phonemes from the middle of a /t/–/d/ continuum that are embedded in lexical items with word-initial /d/, they adapt their existing representation of /d/ to classify the ambiguous tokens as /d/ (Kraljic & Samuel, 2006). This phenomenon has even been seen to impact production (Nielsen, 2014). In both production and perception, the recalibration generalizes to other phoneme pairs that share the same set of features. For example, recalibration on the /t/–/d/ continuum generalizes to the /p/–/b/ continuum. These studies, although focused on monolingual populations, have highlighted the innate connections between groups of phonemes that share similar features. Although the results in the paradigm used in this study cannot be explained by selective adaptation or recalibration, the underlying mechanism may be the same. That is, when L2 learners adapt the VOT properties for one particular phoneme, these innate connections drive change for other phonemes that share the same cues.

Although the mechanisms for L2 phonetic acquisition differ in the segmental and featural approaches, they are not necessarily dichotomous. It is possible that acquisition occurs at both levels in parallel or that a preference for a given approach is dependent on the characteristics of a given sound or contrast. Although the results of this study suggest a featural approach in this particular case, this does not constitute evidence against acquisition at the segmental level for other sounds. Moreover, it is clear that other factors modulate L2 phonetic acquisition, including lexical frequency (Munro & Derwing, 2008) and phonetic context (Munro, Derwing, &

Thomson, 2015), and the interface between such factors and the underlying mechanisms for acquisition warrants further research.

Considering Unit Size and L1 Structure

Within a featural framework to L2 phonetic learning, two issues merit further consideration—unit size and the role of L1 structure. First, although results suggest a featural approach to acquisition for the phonemes of this study, it is worth considering the size or nature of the subphonemic feature. Although previous literature in L2 acquisition has relied on the traditional notion of the contrastive phonological feature (Clements, 1985), there is ongoing debate regarding the minimal unit size relevant for production and perception. As an example, some researchers have interpreted findings from selective adaptation as indicating processing based on acoustic similarity rather than on abstract phonological units (for selective adaptation without phonemic overlap, see Goldinger, Luce, & Pisoni, 1989; for selective adaptation in nonspeech sounds, see Remez, 1979). Another subphonemic approach can be found in word recognition research, with some suggesting that context dependent sublexical units (Reinisch, Wozny, Mitterer, & Holt, 2014) such as allophones (Mitterer, Reinisch, & McQueen, 2018) form the basis for spoken word recognition. Eschewing the traditional fundamental unit debate, Goldinger and Azuma (2003) posit that units may be reconceptualized as self-organizing dynamic states, although again this does not preclude links between different phoneme-like units. Although the results of this study suggest the multisegmental generalization of subphonemic features, the exact mechanisms underpinning these links should be investigated further.

Second, it is worth considering the role of the relationships between the phonetic structures in the L1 and L2. Studies by de Jong, Hao, and Park (2009) and de Jong, Silbert, and Park (2009) focused on L2 learners' abilities to perceive and produce contrasts that are not present in their L1 (see also Brown, 2000). The participants were all L1 Korean learners of L2 English, and these languages differ in their use of nonsibilant fricatives. Thus, results from de Jong and colleagues relate to a learners' ability to learn a new contrast. The study reported here concerned English and Spanish, which employ a bipartite voicing distinction and differentiate between three places of articulation. In contrast to de Jong, Hao, and Park (2009), participants in this study were tasked with adjusting an already existing cue—VOT—in the L1 to approximate L2 targets. Given the close links between the stop consonants in the two languages, it is possible that the existing connections between these consonants in the L1 served as a framework to facilitate the featural acquisition or adjustment.

Exemplifying the role of extant L1 structure, the current results stand in contrast to the production oriented findings of de Jong, Hao, and Park (2009), who found no correlation between the production of the stop–fricative contrast across labials and coronal, leading them to claim that "learners have to acquire two sets of gestures for two places of articulation instead of acquiring one oral gesture that applies to both coronal and labial segments" (p. 369). In the current study, although participants had to coordinate multiple oral gestures for different points of articulation, the gestures themselves existed in the L1. Thus, the crucial task for participants was to reorganize the timing of extant gestures, which may have provided an advantage not available to learners who do not have such a contrast in their L1. De Jong, Hao, and Park (2009) noted that this is a production-oriented effect because previous work in perception had shown correlations for stop–fricative contrast accuracies between different points of articulation. In short, although acquisition may occur at the subphonemic (i.e., featural) level, the ability to accurately produce the contrast and generalize it to other related segments may depend on the available L1 inventory.

Pedagogical Implications *

Although this line of research attempted to leverage phonetic training to better understand the mechanisms involved in L2 phonetic acquisition, several pedagogical implications should be briefly mentioned. First, this study adds to the growing body of work that has demonstrated the effectiveness of visual feedback on L2 phonetic production. Although visual feedback was initially implemented for teaching suprasegmental features (Anderson-Hsieh, 1992; Chun, 1998; de Bot, 1983; Hardison, 2004; Levis & Pickering, 2004), more recently visual feedback in the form of spectrograms and waveforms has been shown to improve segmental production for vowel length (Okuno, 2013), singleton/geminate production (Motohashi-Saigo & Hardison, 2009), vowel formant accuracy (Saito, 2007), and consonantal lenition (Olson, 2014a). More specifically, this study replicated the improvement in VOT previously shown in Offerman and Olson (2016). In addition, the parallel findings for words in isolation and words in connected speech provide further evidence for the notion that training words in isolation may improve production in connected speech (see also Offerman & Olson, 2016; Olson, 2014a).

Second, linking the theoretical and practical implications, this work provides opportunities for enhancing the effectiveness of phonetic instruction. Many current classroom approaches to phonetic instruction tend to adopt a phoneme-by-phoneme approach (see Arteaga, 2000), most notably in lower level language courses. If acquisition is (in part) featural, pedagogical approaches may be adapted to target such multisegment features. That is, instead of training each sound in the L2, instructors may be able to focus on a single feature, or even a phoneme containing the target feature, in order to generate wider ranging improvement. The nature of the specific pedagogical activity (explicit vs. implicit, feature vs. phoneme containing the target feature, etc.) should be empirically tested and compared for effectiveness with other forms of instruction (see Derwing & Munro, 2015, p. 92). Also, although visual feedback may represent an important tool for L2 phonetic instruction, some authors have noted that not all ! sounds, and therefore features, are likely to be so visually intuitive (see Olson, 2014a). Sounds and contrasts that have been successfully addressed with visual feedback, including those that rely on duration, as in this study, may represent a natural starting point.

Pedagogical design aside, the broader implication of these feature-oriented findings is relevant given the generally limited amount of time spent on phonetics in the lower-level L2 classroom (Foote et al., 2011, 2016; Olson, 2014b). Moreover, Lee et al. (2015) found that a longer intervention generally resulted in greater improvement. A refocusing from segment to feature may allow instructors to dedicate greater time to a given feature and thus produce greater improvement across several phonemes.

Conclusion

As learners develop competence in a L2, phonetics can play a key role in establishing comprehensible and intelligible speech, as well as determining the degree of accentedness. A large body of research has demonstrated that learners can develop new phonetic norms in both naturalistic and pedagogical settings, although usually not exactly or reliably nativelike (and nativelikeness is rarely expected or even desired). Although the degree to which learners acquire new phonetic targets varies based on a variety of factors (e.g., age of acquisition, nature of existing L1 network, exposure), there is ongoing debate as to the underlying mechanisms responsible for such acquisition. The results of this study appear to best fit within a featural approach to phonetic acquisition, in which learners acquire subphonemic features (e.g., VOT) that are generalized across multiple phonemes that share the same cue. In addition, although these results point to the acquisition of subphonemic units, it is possible that acquisition takes place across various differently sized units (e.g., features and segments); and while a featural ! approach is possible, it does not preclude segmental acquisition.

The results of this study may serve as the basis for future theoretically and pedagogically oriented research. From a theoretical perspective, although this study supports a featural approach to phonetic acquisition, there are important similarities between the L1 and L2 sounds considered here. Future research should consider the role of extant L1 inventories and their relation to the novel L2 phonemes. In addition, it is acknowledged that the focus on VOT and the use of a single type of phonetic training paradigm is are limiting. Additional research should seek to confirm these results across other features (e.g., voicing) and types of instruction. From a pedagogical perspective, this study investigated VOT, which may influence accentedness in this particular language pairing—L1 English–L2 Spanish—but not intelligibility. Without listener judgments, the practical communicative benefits of the improvement seen here are unclear. This represents a clear pedagogical limitation of this work because intelligibility, rather than a nativelike accent, may be the principal goal of instructors and learners (see Levis, 2005). Although VOT allowed for a gradient analysis of potential featural acquisition, future work should focus on leveraging this approach to improve intelligibility. Moreover, although this line of research is promising, further investigation will be needed to demonstrate how best to incorporate notions of feature acquisition into the L2 language classroom.

Final revised version accepted 24 October 2018

Notes

[1] The Perceptual Assimilation Model–L2 posits that the underlying representation is derived from perceiving “invariants about articulatory gestures” rather than from acoustic information (Best & Tyler, 2007, p. 26).

2 Although feature geometry describes the voicing contrast as a difference in [+/– voice], this ! correlates with the acoustic and articulatory notion of VOT. VOT is generally defined as temporal difference between the release of a stop consonant and the onset of voicing (Lisker & Abramson, 1964). Although VOT is not the sole cue to voicing (e.g., for F0, see Abramson & Lisker, 1985), it is a prominent and reliable cue.

3 The focus here is on general language proficiency or skills courses. General skills courses are defined as those focused on the components of reading, writing, speaking, and listening and generally correspond to beginning and intermediate level courses. General skills courses contrast with those standalone courses (Derwing et al., 2012) that may focus on a particular competence such as pronunciation. An anonymous reviewer noted that it is possible that more advanced courses specifically focused on pronunciation may include a featural approach.

4 Of an original pool of 56 participants, 31 were removed from the analysis for failing to satisfy the inclusionary criteria: six reported speaking a language other than English in the home, 13 reported having taken a phonetics class, and 12 failed to complete all parts of the task.

5 Kartushina et al. (2015), in their review of the literature, differentiated between indirect and direct visual feedback. Indirect visual feedback, the focus of this study, consists of providing raw or abstracted acoustic representations. Direct feedback provides visualization of a participant's articulators (e.g., ultrasound or palatography).

6 Failure to reach nativelike performance was not surprising given that research has demonstrated that longer interventions produce larger gains (Lee et al., 2015), including those using a visual feedback paradigm (Olson, 2014a). However, it is the generalization of such gains from trained to nontrained phonemes that were of import in this study.

7 Analysis was based on 24 participants because isolated word data were missing for one participant.

8 It is not readily apparent why the main experimental and control groups differed in VOT ! production at the pretest. Although all participants fit the inclusion criteria, there was a slight difference in age of acquisition of Spanish between the two groups (*M* = 12.9 years for the experimental group, *M* = 16.8 years for the control group). The matched subgroup was somewhat more comparable with respect to age of acquisition (*M* = 14.1 years).

9 The trend towards VOT reduction for the control group was not entirely unexpected, considering that the same words in isolation were repeated in each session. This effect was not found in the words in connected speech, in which a unique set of target words was recorded each session. Again, although there was a degree of VOT reduction for the control group, this effect was much more pronounced in the experimental group.

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). Orlando, FL: Academic Press.

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, *15*, 517–530. https://doi.org/10.1017/S1366728911000460

Anderson-Hsieh, 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*, 789–797. https://doi.org/10.3758/BF03198414

Arteaga, D. L. (2000). Articulatory phonetics in the first-year Spanish classroom. *The Modern Language Journal*, *84*, 339–354. https://doi.org/10.1111/0026-7902.00073

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using 'Eigen' and S4 (Version 1.1–7) [Computer software]. https://cran.r-project.org/web/packages/lme4/index.html

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, The Netherlands: John Benjamins.

Boersma, P., & Weenink, D. (2017). Praat: Doing Phonetics by Computer (Version 5.1.04) [Computer software]. http://www.praat.org

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal*, *19*, 51–61. https://doi.org/10.18806/tesl.v19i1.919

Brown, C. A. (1997). *Acquisition of segmental structure: Consequences for speech perception and second language acquisition* (Unpublished doctoral dissertation). McGill University, Montreal, Canada.

Brown, C. A. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–64). Hoboken, NJ: Wiley-Blackwell.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229. https://doi.org/10.1006/jpho.1999.0094

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

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, *2*, 225–252. https://doi.org/10.1017/S0952675700000440

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teacher's pronunciation toolbox. In J. Levis & K. Levelle (Eds.), *Social factors in pronunciation acquisition: Proceedings of the 3rdAnnual Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Ames, IA: Iowa State University.

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. https://doi.org/10.1177/002383098302600402

de Jong, K. J., Hao, Y., & Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics*, *37*, 357–373. https://doi.org/10.1016/j.wocn.2009.06.001

de Jong, K. J., Silbert, N. H., & Park, H. (2009). Generalization across segments in second language consonant identification. *Language Learning*, *59*, 1–31. https://doi.org/10.1111/j.1467-9922.2009.00499.x

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, *30*, 22–44. https://doi.org/10.18806/tesl.v30i1.1124

Derwing, T. M., & Munro, M. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research.* Amsterdam, The Netherlands: John Benjamins.

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 599–609. https://doi.org/10.1037/0096-1523.4.4.599

Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 209–220. https://doi.org/10.1037/0096-1523.11.2.209

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99–109. https://doi.org/10.1016/0010-0285(73)90006-6

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*, 70–79. https://doi.org/10.1121/1.396876

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249–289). Philadelphia, PA: John Benjamins.

Flege, J. (1995) Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–73). Baltimore, MD: York Press.

Flege, J. E. (1998) Age of learning and second-language speech. In D. Birdsong (Ed.), *New perspectives on the critical period hypothesis for second language acquisition* (pp. 101–131). Mahwah, NJ: Lawrence Erlbaum.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*, 185–202. https://doi.org/10.1016/0167-6393(87)90025-2 !

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527–552.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English/t/–/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, *92*, 128–143. http://doi.org/10.1121/1.404278

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*, 1–22. https://doi.org/10.18806/tesl.v29i1.1086

Foote, J. A., Trofimovich, P., Collins, L., & Soler Urzúa, F. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, *44*, 181–196. https://doi.org/10.1080/09571736.2013.784345

Green, P., & MacLeod, C. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. https://doi.org/101111/20410210X.12504

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, 305–320. https://doi.org/10.1016/S0095-4470(03)00030-5

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*, 501–518. https://doi.org/10.1016/0749-596X(89)90009-0

Hancin-Bhatt, B. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*, 241–269. https://doi.org/10.1177/026765839401000304

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*, 34–52. https://llt.msu.edu/vol8num1/hardison

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

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure. Typological studies in language* (pp. 229–254). Amsterdam, The Netherlands: 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. *The Journal of the Acoustical Society of America*, *138*(2), 817-832. http://dx.doi.org/10.1121/1.4926561

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*, 262–268. https://doi.org/10.3758/BF03193841

Kuhl, P. K. (1992). Infants' perception and representation of speech: Development of a new theory. In J. Ohala, T. Neary, B. Derwing, M. Hodge, & G. Wiebe (Eds.), *Proceedings of the International Conference on Spoken Language Processing* (pp. 3–10). Edmonton, AL: University of Alberta.

Kuhl, P. K. (1993a). Infant speech perception: A window on psycholinguistic development. *International Journal of Psycholinguistics*, *9*, 33–56.

Kuhl, P. K. (1993b). Innate predispositions and the effects of experience in Speech perception. The native language magnet theory. In B de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht, The Netherlands: Kluwer.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore, MD: York Press.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*, 905–917. https://doi.org/10.1121/1.381770

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*, 345–366. https://doi.org/10.1093/applin/amu040

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

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, *32*, 505–524. https://doi.org/10.1016/j.system.2004.09.009

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

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*, 1–28. https://doi.org/10.1177/002383096701000101

Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, *88*, 557–567. https://doi.org/10.2307/20063159

Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates training with waveform displays. *Language Learning & Technology*, *13*, 29–47. http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf

Morgan, T. (2010). *Sonidos en contexto.* New Haven, CT: Yale University Press.

Munro, M. J. (2016). Pronunciation learning and teaching: What can phonetics research tell us. In T. Isei-Jaakkola (Ed.) *Proceedings of the International Symposium on Applied Phonetics* (pp. 26–29). https://doi.org/10.21437/ISAPh.2016

Munro, M. J., & Derwing, T. M. (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

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics*, *53*, 39–60. https://doi.org/10.1515/iral-2015-0002

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language, and Hearing Research*, *57*, 2065–2075. https://doi.org/10.1044/2014_JSLHR-S-13-0093

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on speech perception: Progress report*, *10*, 357–376.

Mitterer, H., Reinisch, E., & McQueen, J. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, *98*, *77–92.* https://doi.org/10.1016/j.jml.2017.09.005

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

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

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

Olson, D. J. (2014b). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, *97*, 47–68.

R Core Team (2013). R: A Language and Environment for Statistical Computing (Version 3.2.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105. https://doi.org/10.1016/j.wocn.2014.04.002

Remez, R. E. (1979). Adaptation of the category boundary between speech and non-speech: A case against feature detectors. *Cognitive Psychology*, *11*, 38–57. https://doi.org/10.1016/0010-0285(79)90003-3

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research (Version 1.8.4) [Computer software]. https://cran.r-project.org/package=psych

Plonsky, L. & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

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*, 16–40.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, *18*, 452–499. https://doi.org/10.1016/0010-0285(86)90007-1

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*, 177–204. https://doi.org/10.1111/j.1467-1770.1991.tb00683.x

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

Vroomen, J., van Linden, S., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication*, *44*, 55–61.

Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*, 572–577. https://doi.org/10.1016/j.neuropsychologia.2006.01.031

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Appendix S1. Descriptive Statistics by Group for Experiment 1. !

Appendix S2. Linear Mixed-Effects Model for Comparison of Experimental and Control Groups ! in Experiment 2. !

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Learning Sounds in a Second Language Involves Acquiring Generalizable Features

What This Research Was About And Why It Is Important

In learning a second language, adults must not only learn new words and grammatical structures, they also must learn a new system of sounds to be able to correctly pronounce the new language. Previous research on how learners acquire new sounds has generally taken a "segmental" approach, meaning that learners acquire the new system on a sound-by-sound basis. This approach is also common in the language classroom, in which both textbooks and instructors generally teach pronunciation for individual sounds (e.g., consonants or vowels). However, some researchers have suggested that learning might take place at the "feature" level, with each individual sound being composed of several components or features such as the presence or absence of voicing or a specific place of articulation. Some sounds may have one of more of these features in common. This study investigated how new sounds are acquired, and whether there is evidence for adult learners acquiring individual features. The researcher found that if learners improved their pronunciation of the sound they were trained on, their pronunciation also improved for those sounds that were not targeted in training but that share a similar feature.

What the Researchers Did

- The researcher tested 25 adult English-speaking learners of Spanish in a training study targeting one of three Spanish consonants ("p," "t," or "k") that share the same feature of voicing. In English, these sounds are followed by a long puff of air (aspiration), but in Spanish they have much less aspiration.

- Training consisted of learners recording themselves speaking Spanish and examining a visual representation of the target sounds (sound wave) in a speech editing program. The learners also compared the images of their productions in a speech editing program with the sound wave produced by a native Spanish speaker. After comparison, the learners rerecorded themselves.
- The learners' productions were recorded before instruction (pretest), immediately after instruction (posttest), and four weeks after instruction (delayed posttest). !
- Analysis compared how the learners improved in their pronunciation of both the sound that they studied (e.g., "p") and the other related sounds that they did not explicitly study (e.g., "t" and "k"). An improvement would be demonstrated if learners produced more Spanish-like consonants characterized by shorter aspiration.

What the Researchers Found

- The learners' pronunciation of the target sounds improved following the training, that is, when the learners were trained on "p," they also improved in their pronunciation of this consonant.
- The learners' pronunciation also improved for the other related sounds, even though they were not addressed by the visual comparison activity. In other words, the learners who were trained on "p" also improved in their pronunciation of "t" and "k."
- The amount that each learner improved on the trained and not trained consonants was highly related. Learners that improved a lot on the trained consonants also improved a lot on the other related consonants.

Things to Consider

- In this study, receiving training on one consonant (e.g., “p”) led to improvements on other related consonants (e.g., “t” and “k”). This suggests that learning might not happen on a sound-by-sound basis. Some learning might happen at the feature level.
- Although this study shows that some learning happens at the feature level, it is possible that different sounds are learned at different levels—sound-by-sound, feature-by-feature, or a combination of the two.
- Instructors might want to consider if it is more efficient for them to teach each sound of the new language individually, or if they can target one feature to improve multiple sounds.

How to cite this summary: Olson, D. J. (2018). Learning sounds in a second language involves acquiring generalizable features. *OASIS Summary* of Olson in *Language Learning*. https://oasis-database.org