Journal of Phonetics 52 (2015) 105–120

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Effects of perceptual phonetic training on the perception and production of second language syllable structure

Amanda Huensch[a,*], Annie Tremblay[b]

[a] University of South Florida, 4202 E. Fowler Ave., CPR 419, Tampa, FL 33620-5700, USA
[b] University of Kansas, 1541 Lilac Lane, Blake Hall, Room 427, Lawrence, KS 60045-3129, USA

A R T I C L E I N F O

Article history:
Received 29 September 2014
Received in revised form
4 June 2015
Accepted 27 June 2015
Available online 24 July 2015

Keywords:
Speech perception
Speech production
Phonetic training
Syllable structure

A B S T R A C T

This study investigated the effect of perceptual training on second language (L2) learners' perception and production of syllable structure, thereby shedding light on the relationship between L2 speech perception and production and on the nature of stored representations. Korean L2 learners of English completed perceptual training on palatal codas in a pretest–post-test design. We compared the effects of training on improvements in perception and production for trained and new words and talkers. A control group who completed an unrelated perceptual training was included for comparison. Results indicated that learners who received perceptual training on palatal codas outperformed those who did not in perception and production tasks and generalized learning to new words and new talkers. Yet perception improvements were not directly linked to production improvements. The finding that perceptual training improved production and allowed for generalizability to new words and talkers in both perception and production provides evidence that L2 perception and production systems are linked. However, the lack of a one-to-one relationship between perception and production improvements suggests that the representations underlying L2 speech perception and production may be distinct.

1. Introduction

One central question in the field of second language (L2) phonology is the role that speech perception plays in speech production and the degree to which the speech perception and production systems are linked. Some research suggests that L2 learners' speech production errors may stem from perceptual errors (e.g., Brannen, 2002; Broselow, 2009; Flege, 1995; Flege, MacKay, & Meador, 1999; Peperkamp & Dupoux, 2003; Rochet, 1995), whereas other research suggests that speech production errors are also driven by the non-native-like timing of gestures (e.g., Colantoni & Steele, 2008; Davidson, 2010; Davidson & Stone, 2003; Oh, 2008; Zsiga, 2003). Ultimately, while it is likely that both perceptual and articulatory problems contribute to learner production errors, what remains unclear is the degree to which L2 learners' perception and production systems are linked. The answer to this question depends, at least in part, on the theory of speech perception that is presumed.

On the one hand, psychoacoustic theories of speech perception assume the primitives of speech perception to be the acoustic properties of the speech signal (e.g., Diehl, 1987; Diehl & Kluender, 1989; Holt, Lotto, & Kluender, 1998; Kingston & Diehl, 1994; Kluender, 1994; Kuhl, 1985; Lotto, 2000; Neary, 1990; Ohala, 1996). The Speech Learning Model (SLM) (Flege, 1991, 1995, 2003) falls within this theoretical framework and is concerned with the difficulties learners have in the ultimate attainment of L2 segments. It posits that when learning a native language (L1), a child becomes attuned to the sound contrasts in that language and stores language-specific aspects of these sounds to form phonetic categories (based on acoustic properties of the speech signal) in long-term memory. However, these categories are not fixed and can change over time. The SLM explicitly postulates that not having accurate perception will lead to problems in production. However, because the SLM assumes a psychoacoustic view of speech perception, it does not predict a one-to-one relationship between perception and production, nor does it predict that all production errors are perceptually based. In other words, although perceptual and production systems may be indirectly linked, they do not share representations.

* Corresponding author. Tel.: +1 813 974 2548; fax: +1 813 905 9937.
E-mail addresses: huensch@usf.edu (A. Huensch), atrembla@ku.edu (A. Tremblay).

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2015.06.007

Direct Realist theories of speech perception (Fowler, 1986, 1989, 1991, 1994, 1996), on the other hand, take articulatory gestures as the direct primitives of speech perception. Within this theoretical framework, the Perceptual Assimilation Model (PAM) (Best, 1994, 1995; Best, McRoberts, & Goodell, 2001) posits that non-native speech sounds will be perceived in relation to their gestural similarities to and differences from L1 speech sounds. Non-native segments can be (1) assimilated to a native category (as either a good, acceptable, or noticeably deviant exemplar), (2) perceived as an uncategorizable speech sound, or (3) not perceived as a speech sound (Best, 1995). PAM was originally proposed to account for cross-linguistic speech perception in naïve listeners, but the model was extended to the domain of L2 speech learning (PAM-L2; Best & Tyler, 2007): If a non-native sound assimilates to a native sound because it is a good phonetic exemplar of that sound (i.e., (1)), the learning of the non-native sound is likely to be small. However, the authors note that within a Direct Realist framework, learning continues into adulthood, so it is possible that categories could shift. According to Best and Tyler, the likelihood that a category would shift depends to a large degree on whether L2 learners perceive a difference between L1 and L2 sounds. They hypothesize that communicative pressure may further increase the likelihood of a category shift if the assimilation leads to confusion between high-frequency words in phonologically dense neighborhoods. Best and Tyler emphasize that PAM-L2 differs from the SLM in the perceptual primitives of speech perception it assumes, with these primitives being distal articulatory gestures rather than proximal acoustic cues. Moreover, Best and Tyler highlight that PAM-L2 differs from the SLM in the importance it places on phonetic *and* phonological categories (unlike the SLM, which places importance only on phonetic categories). Ultimately, because PAM and PAM-L2 have their roots in Direct Realism, they posit linked perception and production systems that share representations, and would thus predict that perception and production learning would be strongly correlated.

The present study investigates the nature of the relationship between L2 learners' perception and production systems with the goal of shedding further light on whether a model couched within a psychoacoustic theory of speech perception (e.g., the SLM; Flege, 1991, 1995, 2003) or one couched within a Direct Realist theory of speech perception (e.g., PAM(-L2); Best, 1994, 1995; Best et al., 2001; Best & Tyler, 2007) can better account for L2 speech learning. The hypotheses of the SLM and PAM(-L2) have been tested in many studies, yet sometimes without consideration for the larger theoretical framework within which they are couched, a valid point also made by Best and Tyler (2007). Since neither model focuses explicitly on the nature of the relationship between L2 learners' speech perception and speech production, the present research may speak more about these broader theoretical frameworks than about the models themselves. However, it also emphasizes the need to carefully consider these theoretical frameworks when choosing a particular L2 speech learning model.

In this study, we investigate the mediating effect of perceptual multi-talker training (also referred to as high-variability phonetic training) on L2 learners' speech perception and speech production. Because such training has been shown to alter perceptual representations (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Iverson & Evans, 2007; Iverson, Hazan, & Bannister, 2005; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991), it can provide a critical window for examining how speech perception affects speech production. A model couched within a psychoacoustic view of speech perception (e.g., the SLM) might not predict a direct link between L2 learners' perceptual learning and production improvements following such training, even in the presence of overall production improvements; by contrast, a model couched within a Direct Realist view of speech perception (e.g., PAM(-L2)) would predict that perceptual learning should directly translate into production improvements.

Unlike previous research on perceptual multi-talker training, the present study focuses on L2 learners' perception and production of syllable structure. As such, it accomplishes an additional goal: to determine whether the existing benefits of perceptual multi-talker training shown with individual segments (e.g., Bradlow et al., 1997; Iverson & Evans, 2007; Iverson et al., 2005; Lively et al., 1993; Logan et al., 1991) can extend to the perception and production of L2 syllable structure. Korean speakers with a range of proficiency levels in English completed pretests and post-tests eliciting the perception and production of palatal codas in English. Since Korean is far more limited than English in the segments it allows in coda position, it is a well-suited L1 for investigating the perception and production of L2 syllable structure.

2. Relationship between L2 perception and L2 production

A large body of research has investigated the relationship between speech perception and speech production in L2 learners, focusing largely on individual segments. Some of this research has shown that non-native speakers may learn to produce L2 segments accurately without being able to perceive them accurately. One well-known example is the learning of the English /ɹ/-/l/ contrast by native Japanese speakers. Goto (1971), and later Sheldon and Strange (1982), showed that Japanese L2 learners of English could produce the English /ɹ/-/l/ contrast accurately despite having poor perception of it. These findings were attributed in part to the fact that Japanese L2 learners of English often receive articulatory training that eventually enables them to produce the English /ɹ/-/l/ contrast accurately (e.g., see also Flege, Takagi, & Mann, 1995); however, they never quite develop the ability to map the acoustic cues of English /ɹ/-/l/ onto separate phonetic categories, a topic that continues to be the object of much research (e.g., Ingvalson, Holt, & McClelland, 2011; Ingvalson, McClelland, & Holt, 2011; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). Similar findings were reported in studies on other segments (e.g., German speakers' production and perception of English /ɛ/-/æ/; Bohn & Flege, 1990).

However, perhaps the majority of L2 studies that elicited both speech perception and speech production data have instead suggested that accurate perception may be a necessary precursor of accurate production (without entailing it). For example, Rochet (1995) demonstrated that Portuguese L2 learners of French more often produced /y/ as an /i/-like vowel, whereas English speakers more often produced /y/ as an /u/-like vowel; these results paralleled findings from his perception task, where /y/ vowels were categorized more often as /i/ by Portuguese speakers and more often as /u/ by English speakers. Rochet thus claimed that the accented speech of these learners might stem from their inaccurate perception. Likewise, Flege, Bohn, and Jang (1997) found that L2 learners' perceptual sensitivity to spectral and duration information in the English vowels /i ɪ ɛ æ/ could account for a significant amount of variance in L2 learners' productions. However, the authors also point out that a large amount of variability in production data was not accounted for by the perceptual data. Therefore, while there were some indications of a relationship, there was no compelling evidence for a direct link. Similarly, Flege et al. (1999) investigated the perception and production of 10 English vowels by Italian L2 learners of English. They found that L2 learners' vowel production accuracy (reflected in intelligibility scores obtained from native English speakers) correlated significantly with their vowel discrimination scores. However, they also noted that the learners' discrimination scores did not account for a large amount of variance in their production intelligibility, suggesting that other factors were at play in explaining L2 learners' production accuracy.

More recently, Evans and Iverson (2007) examined the relationship between speech perception and speech production, but in speakers of British English who spoke different English dialects and had different regional accents. In a longitudinal study, they investigated the perception and production of 11 English vowels of individuals from Northern England who attended university in the south of England. Results from perception and production tasks administered at four separate times were compared to determine whether perceptual changes were reflected in production. The group results indicated that the participants' vowel productions changed, but their perceptions did not change. Despite this finding, Evans and Iverson provided some evidence for a link between perception and production systems: Those who produced more southern vowels rated southern vowels as better in the perception task, and those who produced more northern vowels rated northern vowels as better in the perception task. These findings provide evidence outside L2 acquisition research for a connection between perception and production. Nevertheless, similar to findings from L2 research, this relationship does not appear to be direct.

A more robust test of the relationship between perception and production would instead be to investigate how learning in one skill (e.g., perception) affects learning in the other (e.g., production). One method for doing this involves perceptual phonetic training. The type of training upon which perceptual training in this research is based is called high-variability phonetic training (HVPT). The use of perceptual training was chosen because beyond allowing learners to establish new categories, this training has the advantage of providing a window into the relationship between perception and production by modifying L2 learners' perceptual representations. In their widely cited studies, Logan et al. (1991) and Lively et al. (1993) put forth a highly effective perceptual training method that improved on previous methods by training learners with multiple words from multiple talkers (between 4 and 6 talkers, as opposed to one). L2 learners who are exposed to this type of training should establish more robust categories and thus be able to generalize learning to new words and new talkers. The HVPT method generally includes a pretest–post-test design with a few weeks of perceptual training in between. In Logan et al. and Lively et al., the pretest, perceptual training, and post-test all used a forced-choice word-identification task.

Important to the question of the relationship between perception and production is whether perceptual training can lead to enhancement in L2 speech production. In a continuation of Logan et al. and Lively et al., Bradlow et al. (1997) investigated the effects of perceptual training on the production of /ɹ/ and /l/ in Japanese L2 learners of English. A comparison of trained learners and untrained learners revealed that the trained learners' perception scores (accuracy) and production scores (goodness judgments from native English listeners) were significantly higher than those of the untrained learners (for similar group results, see Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005). However, in their comparison of individual perception and production scores, Bradlow et al. found substantial variability in production gains among the trained learners and no correlation between perception and production gains. One explanation Bradlow et al. (1997) offer for their findings is that different participants may learn the required motor commands at different rates, with not all participants who modified their underlying perceptual representation necessarily making the corresponding changes in speech production. Hence, it may be the case that the timing of the articulatory commands is somewhat independent of the creation of perceptual representations.

This lack of a direct relationship between perception and production gains may suggest that the representations underlying speech perception differ, at least to some extent, from those underlying speech production. PAM(-L2) posits linked systems that share representations, and thus while ultimately a perception theory, it would predict that perception and production learning would be strongly correlated. The SLM makes strong claims about perception leading production. However, it posits an indirect relationship between modalities, with perceptual learning allowing for a reorganization of the acoustic-auditory space that ultimately feeds the system used for both perception and production. Because it assumes a psychoacoustic view of speech perception, the SLM would not necessarily predict that perception and production learning would be strongly correlated.

The present study aims to shed further light on the nature of the relationship between L2 speech perception and L2 speech production by examining the effect of perceptual training on both. However, unlike previous studies, it does so by examining how training affects the perception and production of syllable structure.

3. L2 perception and L2 production of syllable structure

Restrictions on the types of segments that syllables can have in the L1 have been shown to affect both L2 speech perception (e.g., Davidson, 2011; Davidson & Shaw, 2012; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kaheki, & Mehler, 2001; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak & Idsardi, 2007) and L2 speech production (e.g., Broselow, Chen, & Wang, 1998; Broselow & Finer, 1991; Carlisle, 1997, 1998; Davidson, 2006, 2010; Eckman & Iverson, 1993; Hancin-Bhatt, 2000; Hancin-Bhatt & Bhatt, 1997).

Using a nonword vowel perception task, Dupoux et al. (1999) showed that native speakers of Japanese, a language with a limited syllable structure inventory, were much more likely to hear an “illusory” vowel between two word-medial voiced obstruents (e.g., *ebzo*) than native speakers of French, a language with a much broader syllable structure inventory. Dehaene-Lambertz et al. (2000) replicated these findings using evoked potentials in an oddball paradigm, and Dupoux et al. (2001) demonstrated that this illusory vowel effect was prelexical rather than lexical. Kabak and Idsardi (2007) further demonstrated that it was syllable structure violation rather than consonantal contact that was more likely to cause an illusory vowel effect in Korean L2 learners of English (see also Davidson, 2011).

A more recent study by Dupoux et al. (2011), however, suggests that the illusory vowel effect might instead stem from L1-specific phonetic cues. Using both a vowel-detection experiment and an ABX discrimination experiment, the researchers showed that Japanese and Brazilian Portuguese listeners were more likely to hear an illusory vowel between word-medial consecutive stops than were European Portuguese listeners. Most of the stimuli in their experiments contained sequences of voiced stops, which can occur at the surface level in European Portuguese (via vowel deletion) but not in Japanese or Brazilian Portuguese. These results thus suggest that it is the surface occurrence of sound sequences (as delimited by syllable boundaries) that appear to determine the perception of an illusory vowel.

Further evidence for the influence of L1-specific phonetic cues comes from Davidson and Shaw (2012), who examined whether native English listeners' perception of word-initial consonant clusters would vary as a function of the repair form (e.g., vowel epenthesis or prosthesis, consonant deletion or substitution) of the comparison stimulus. Across two discrimination tasks, the authors found that perception was more difficult for stop-stop sequences when the comparison stimulus contained an epenthetic vowel than when it contained other repair forms. They attributed this finding to the fact that in English words, bursts occurring between a word-initial stop and a subsequent stop are indicative of a reduced vowel (e.g., *begin, potato*). The results also showed that perception was more difficult for stop-nasal sequences when the comparison stimulus contained a consonant deletion or substitution than when it contained other repair forms. They attributed this finding to the lower intensity of stop bursts before nasal consonants, which made the identity of the stop less easily identifiable for English listeners. Davidson and Shaw's (2012) findings suggest that a successful account of the illusory vowel effect ought to consider the influence of L1-specific phonetic cues.

Syllable structure can pose difficulties for non-native listeners not only at the level of speech perception, but also at the level of speech production. Some researchers have proposed that L2 learners have more difficulty producing sequences of sounds that are closer together in the Sonority Hierarchy (e.g., Broselow & Finer, 1991; Eckman & Iverson, 1993; for a similar account of illusory vowels in speech perception, see Berent, Steriade, Lennertz, & Vaknin, 2007). Other researchers have attributed L2 learners' adoption of different strategies to repair illegal sound sequences to L1 syllable structure constraints (e.g., Broselow et al., 1998; Carlisle, 1997, 1998; Hancin-Bhatt, 2000; Hancin-Bhatt & Bhatt, 1997).

Other researchers, however, have instead suggested that L2 learners' production of illusory vowels is instead the result of the inaccurate timing of overlapping articulatory gestures, with these vowels thus differing qualitatively from those inserted for the purpose of phonological repair. Using a word repetition task, Davidson (2006) showed that English speakers with no knowledge of Slavic languages produced vocalic elements in Czech fricative-stop sequences acoustically differently from the schwa they produced in Czech fricative-schwa-stop sequences (see Davidson, 2005, for ultra-sound evidence that these were also articulated differently). Based on these results, Davidson argued that the vocalic element in the English speakers' fricative-stop sequences was not a true epenthetic schwa, but rather the result of gestural mistiming. Davidson (2010) replicated these results in a later study where English speakers both heard and read the words to be produced on a computer screen, suggesting that the transitional schwa that English speakers produced could not be attributed to perceptual difficulties.

The present study focuses on the relationship between L2 learners' perception of illusory vowels and their production of such vowels. Assuming that the intrusive vowels that L2 learners produce may be due to their mistiming of articulatory gestures (following Davidson's work), if L2 learners' speech perception is not rooted in such gestures (as proposed by psychoacoustic theories of speech perception such as the SLM), we may expect that L2 learners will not show a direct relationship between their perception and production of syllable structure. A direct relationship between the two would instead suggest that both L2 speech perception and L2 speech production are rooted in the same gestural representations (as proposed by Direct Realist theories of speech perception such as PAM(-L2)). We examine this relationship by investigating the effect of perceptual training on Korean speakers' perception and production of coda consonants in English.

Differences between Korean and English provide an interesting test case for investigating questions concerning the relationship between perception and production and the role of syllable structure restrictions. Korean and English both allow coda consonants, but they differ in the segments allowed in that position. More specifically, English contains obstruents in the following categories: six plosives /p b t d k g/, nine fricatives /f v θ ð s z ʃ ʒ h/, and two affricates /ʧ ʤ/, and allows for a variety of syllable types: (C)(C)(C)V(C)(C)(C)(C). Importantly, English allows palatals in coda position.[1] On the other hand, the Korean obstruents consist of nine stops (unaspirated/lenis /p t k/, aspirated /pʰ tʰ kʰ/, and fortis /p' t' k'/), three fricatives (/s s' h/), and three affricates (/ʧʰ ʧ' ʧ/). Korean allows V, CV, and CVC syllables, but has a robust system of neutralization in codas, where stops neutralize to the lenis variety. Only lenis stops, nasals /m n ŋ/, and the lateral /l/ are allowed in codas (Yeon, 2004). Korean contains /ʃ/, but only as an allophone of /s/ before high vowels.[2] Investigations of Korean adaptations of loanwords show that palatals in word-final codas are produced with an epenthetic [i] (e.g., Kim, 2009). When producing English words, Korean L2 learners of English sometimes produce an epenthetic

[1] Some dialects of English do not have /ʒ/ in final position whereas others do. These differences emerge in words like *garage*.

[2] Note that since the present study does not tease apart syllable structure restrictions from L1-specific phonetic cues (cf. Davidson, 2011; Davidson & Shaw, 2012; Dupoux et al., 2011), we remain agnostic as to whether the nature of the L1 influence investigated in this study is phonetic (i.e., the result of particular phonetic cues that arise from sound sequences in

vowel in coda or word-final position following English palatals (both fricatives and affricates, /ʃ ʧ ʤ/), resulting in productions such as *language*[i] instead of *language* (Schmidt & Meyer, 1995).

Using a pretest–post-test design, we examined whether perceptual phonetic training on palatal codas has an effect on the speech perception and production of Korean L2 learners of English, whether improvements generalized to new words and new talkers, and whether perception gains co-varied with production gains. The results will allow us to determine whether perceptual phonetic training can successfully extend to syllable structure while shedding further light on the relationship between perception and production and thus on the nature of perceptual representations.

4. Methods

4.1. Participants

Participants included 24 adult Korean L2 learners of English randomly assigned to two groups. The experimental group ($n=12$, 5 women) received perceptual training on palatal codas, and the control group ($n=12$, 5 women) received perceptual training on vowel pairs (no words in the control training contained palatal codas). This latter group completed both perception and production pretests, a training task unrelated to palatals to ensure a similar amount of time on task, as well as the post-tests. Five native English listeners (3 women) also completed the perception pretest.

Participants completed a cloze test as a measure of their global proficiency in English (for discussion, see Tremblay, 2011) and a language background questionnaire. Table 1 shows the means and ranges for the cloze test scores and a subset of relevant language background information.

In order to verify that the experimental and control groups were not different prior to training, a series of one-way ANOVAs were conducted comparing the two groups on each of the variables in Table 1. No significant differences were found between the two groups on any of these measures, $F<1$.

4.2. Materials

In the pre- and post-test phases, participants completed a forced-choice word-identification experiment and a read-aloud production experiment in which they read the words and sentences that had been heard in the forced-choice word-identification task. In the perceptual training phase, participants in the experimental group completed a forced-choice word-identification task.

Experimental stimuli were 48 minimal pairs of natural tokens, half of which were real words and half of which were nonce words (see the Appendix for a complete list of stimuli). Each pair was comprised of a monosyllabic word ending with a singleton palatal in coda position (CVC) and the adjectival derivation of the same word (CVC*y*) (e.g., real words: *push*/*pushy*, *dodge*/*dodgy*, *catch*/*catchy*; nonce words: *mish*/*mishy*, *tudge/tudgy, tetch/tetchy*). The stimuli included both real and nonce words because of the limited availability of such word pairs in English and to minimize potential word frequency effects in the task. Twenty-four pairs (8 with each palatal consonant) were included both in the experimental training and in the pretests and post-tests, and the remaining 24 pairs (8 with each palatal consonant) were included only in the pretests and post-tests. Thus, only half of the stimuli were presented in the experimental training. This was done to determine whether learners can generalize improvement from words in the training ('trained' words) to words not in the training ('new' words). The words not included in the training were in both the pretests and post-tests. Hence, if the learners who completed the experimental training perform significantly less accurately on the new words than on the trained words, this difference will not be attributable to practice effects.

In both the perception tests and the experimental training, minimal pairs were presented in isolation as well as within the carrier sentences *He said* ___ *angrily* and *He said* ___ *frequently*. These carrier sentences provided a context in which the target word is followed by a vowel (*angrily*) or a consonant (*frequently*). A sentential context was provided in case the learners' performance on words in isolation would be at ceiling. Having both prevocalic and pre-consonantal conditions would also allow us to provide more phonetic variability in the training contexts. The 48 stimuli were thus presented three times in the pretest and three times in the post-test: for each test, they occurred twice in context (*angrily*/*frequently*) and once in isolation. Since it was the case that the isolated words context appeared to be relatively easier for the learners, below we report only the results for words in sentences. Following the design of Bradlow et al. (1997), there were also 28 minimal pairs that contrasted other phonemes of English, both in isolation and within sentences.

All of the stimuli were recorded by six native speakers of English (3 women). Two talkers were from the Inland North, three were from the Midland, and one was from the South (Labov, Ash, & Boberg, 2006). All talkers had been living in the (Midwestern) city where data collection took place for at least 4.2 years at the time of recording. Stimuli were recorded in a sound-attenuated booth via a Marantz PMD570 solid state recorder using either an Earthworks M30 standing microphone or an AKG c520 head-mounted microphone at 44.1 kHz. Stimuli were then segmented into individual files and normalized to 65 dB in Praat (Boersma & Weenink, 2010). Recordings from two men and two women were used in both the pretests and post-tests as well as in the training ('trained'

(*footnote continued*)
particular syllable positions) or phonological (i.e., the result of more abstract L1 syllable structure restrictions). In light of the research that suggests this L1 effect is delimited by syllable boundaries (Kabak & Idsardi, 2007), we continue to refer to it as an effect of syllable structure (rather than phonotactic) restrictions.

Table 1
Language background information of participants.

		Cloze test (/50)	Age	Age of first exposure to English (years)	Years of instruction	Years of residence in English-speaking environment	Daily % use of English
Native English (*n*=5)	Mean	n/a	27	From birth	n/a	n/a	97
	SD	n/a	7.3	n/a	n/a	n/a	4.0
	Range	n/a	22–39	n/a	n/a	n/a	90–100
Experimental group (*n*=12)	Mean	28	30	11	10	4.1	42
	SD	7.4	8.7	4.3	5.7	3.5	27.3
	Range	9–38	18–46	0–15	3–22	0–10	5–85
Control group (*n*=12)	Mean	27	30	12	9	4.5	49
	SD	10.1	8.4	2.7	3.7	4.0	35.3
	Range	15–38	21–48	6–16	5–17	0.2–13.3	10–99

talkers), while the recordings from the other man and woman were used only in the pretests and post-tests ('new' talkers). The two talkers not used for training were chosen at random.

4.3. Procedures

The procedures consisted of a pretest phase, a perceptual training phase, and a post-test phase conducted over approximately 10 days. The pretests and post-tests were administered individually in a quiet room. The perceptual training phase was completed online.

4.3.1. Pretest phase

The pretest phase consisted of both perception and production experiments. The perception tests included a forced-choice word-identification task of words in isolation and words in carrier phrases. Stimuli were presented using E-Prime (Schneider, Eschman, & Zuccolotto, 2001a, 2001b), and participants wore either Beyerdynamic DT 770 or Sony MDR 7506 headphones and had control over the volume level via an Alesis iO2 USB interface. At the beginning of each trial, participants heard a word or sentence and then saw the two words or sentences from each pair presented on the left and right side of the screen. They were instructed to choose the correct response as quickly as possible by pressing one of two marked keys on the keyboard.

Each of the 96 experimental words (from the 48 minimal pairs) and 56 filler words (from the 28 minimal pairs) were presented once in isolation (in one block) and once in each carrier-phrase context (in another block), along with eight practice items at the beginning of each block to familiarize participants with the procedure. The pretest thus included a total of 472 trials. The isolated-word block lasted approximately 12–15 min and the carrier-phrase block lasted approximately 27–32 min. Participants were offered breaks after both the isolated-word block and the carrier-phrase block and at one third and two thirds of the way through the carrier-phrase block. Whether the correct word was on the left or right was counterbalanced across trials. The talker that produced the stimuli was also counterbalanced across trials, such that a learner did not hear both words from a minimal pair spoken by the same talker. Six lists were created to counterbalance across participants, such that all words from all talkers were heard across different participants. The order of block (whether a participant began with words in isolation or words in a carrier phrase) was counterbalanced across participants. Five native English listeners also completed the perception pretest.

The production pretest was completed after the perception pretest and included a read-aloud task modeled on the perception task. Recordings were completed in a sound-attenuated booth via a Marantz PMD570 solid state recorder using an AKG c520 head-mounted microphone. The first set of participants was recorded at 44.1 kHz, but the settings were changed and the remaining participants were recorded at 48 kHz. However, recordings at 48 kHz were all converted to 44.1 kHz for the assessment phase. All participants' pretest and post-test productions of target words in context (*angrily*/*frequently*) and in isolation were segmented into individual files and normalized to 65 dB in Praat (Boersma & Weenink, 2010) for later analyses.

In the read-aloud task modeled on the perception task, participants received a visual word or sentence prompt and read the word or sentence. All stimuli (real and nonce words in isolation and in sentences) were combined and manually randomized. Participants were instructed to read at a comfortable pace and to give their 'best guesses' for any unfamiliar words. Participants recorded all 456 tokens and were offered a break one third and two thirds of the way through this task.[3] The duration of the task was approximately 15–30 min, depending on the reading pace of participants and whether they took breaks.

4.3.2. Experimental training phase

The perceptual (palatal coda) training phase for the experimental group consisted of eight, 20-min daily sessions of online training delivered via Paradigm Player (Perception Research Systems, 2007). Each training session consisted of a forced-choice word-identification task with two blocks: one with words in isolation and the other with words in carrier phrases. Participants always began

[3] In some cases, the participant did not record the word. From a total of 10,944 possible words (456 productions × 24 speakers), this occurred 16 times. In these cases, the items were omitted from the analysis.

with words in isolation and continued with words in carrier phrases. Each block consisted of (a) the set of 48 words in isolation from one talker along with 16 fillers, or (b) the set of 96 words in each of the carrier contexts from one talker along with 32 fillers. During each session day, learners heard stimuli from two different talkers (of the four who were randomly selected to be training stimuli). Blocks were counterbalanced such that over the course of the eight sessions, learners heard each word in isolation and in each carrier phrase twice from each of the four talkers.

The procedures for each trial were identical to the word-identification task of the perception pretest and post-test, except that (a) during training, participants received feedback as to whether or not they answered correctly, and (b) participants saw the words for 500 ms before the audio stimulus. For every response (whether correct or incorrect), participants heard the stimulus again during the feedback screen. During each training day, participants spent approximately 20 min on task, for a total of approximately 160 min of training.

Participants were instructed to begin the perceptual training sessions as soon as possible, but no earlier than the day following the pretest. They were also instructed to complete the sessions in eight successive days, with a night's sleep in between each session. For the purpose of learning, it is important that participants wait at least one night before doing the next session, because the brain consolidates information while asleep (e.g., Marshall & Born, 2007; Stickgold, 2005; Walker & Stickgold, 2004). No participant completed two sessions in the same day. Nevertheless, because of participants' schedules, sometimes there was more than one day between sessions.

4.3.3. Control training phase

The perceptual training for the control group consisted of eight daily sessions of online training delivered via Pierceive. This online tool was developed at the home institution to deliver perceptual training for another study conducted by another researcher. The focus of the perceptual training for the control group was on three vowel pairs (/æ/-/ɛ/, /i/-/ɪ/, and /oʊ/-/ʊ/), all presented in monosyllabic nonce words. The stimuli contained ten consonants (/d t n b p m k g h s/) in onset and/or coda position. None of the stimuli contained palatal consonants. The nonce words were always heard in isolation. Talkers were eight native speakers (4 women) of North American English. Their speech rate varied from slow/careful to normal/casual to fast. Participants were randomly assigned to one of four training conditions within which stimuli varied along the consonant context, talker, and speech rate dimensions.

In each training session, participants heard the word pairs in a self-paced exposure phase and were tested on the word pairs in a subsequent testing phase. Participants were instructed to spend approximately 20 min on the exposure phase to mirror the time spent by the experimental group; however, unlike in the experimental perceptual training, these participants controlled the amount of time spent on task. Because of this, time on task varied across participants. The testing phase in the control training used a forced-choice word-identification task, as did the experimental training.

The experimental and control groups were similar in terms of: (1) days between the pretest and start of training (experimental group: 3.5 [SD = 2.4]; control group: 3.3 [SD = 3.9]); (2) days off during training (experimental group: 1.6 [SD = 1.0]; control group: 2.3 [SD = 1.5]); and (3) days between the end of training and the post-test (experimental group: 1.5 [SD = 0.8]; control group: 1.3 [SD = 0.6]). A series of one-way ANOVAs indicated no significant differences between the groups on any of these measures, $F<1$ for (1) and (3), $F(1,22)=1.67$; $p<0.210$ for (2). In contrast, because the control group did not adhere to the instructions of spending 20 min per day on training, the total time on task was quite different between groups (experimental group: 160 min; control group: 68 min [SD = 26]). Ultimately, the control group spent significantly less time completing perceptual training than the experimental group, $F(1,22)=150.64$; $p<0.001$. We will return to this point in Section 5.1 when we present the perception results.

4.3.4. Post-test phase

The post-test phase was identical to the pretest phase, including both the forced-choice word-identification tasks and the production task. The perception tests were balanced across participants such that if participants began with the words in isolation in the pretest, they began with the words in carrier phrases in the post-test, and vice versa. The production task was again completed after the perception task.

4.4. Data analysis

Answers on the perception tests were scored as either accurate or inaccurate. Participants' results on the perception task were then transformed into d' scores. Calculating d' scores is a method used within Signal Detection Theory to provide a measure of listeners' sensitivity (Macmillan & Creelman, 1991). The d' calculation is done by converting proportions of *hits* (H) (i.e., identifying A as A) and *false alarms* (F) (i.e., identifying B as A) into *z*-scores under a normal distribution: $d'=z(\text{H})-z(\text{F})$. In the current experiment, the potential bias would be the likelihood of choosing the disyllabic word (e.g., *pushy*) over the monosyllabic one (e.g., *push*).

Production responses were rated by a group of native English listeners (NLs) in two listening tasks: a paired-comparison task and a forced-choice word-identification task. NLs were 67 (18 men and 49 women; mean age: 22, SD: 6.6, range: 18–61) native speakers of English who had learned only English between the ages of 0–5. Most of them were undergraduate students at a Midwestern university. In addition to the production assessment tasks, each listener filled out a language background questionnaire. NLs' mean reported daily use of English was 98% (SD: 4.0, range: 80–100). Of the 67 NLs, 43 completed both the forced-choice word-identification and paired-comparison tasks, but never with productions from the same learner. Of the 43 listeners who completed both tasks, 15 did so with approximately one week in between tasks. The other 28 listeners completed both tasks on the same day. The 43 listeners who completed both tasks were balanced for whether they started with the paired-comparison task or the forced-choice word-identification task. The remaining 24 listeners completed only one of the two tasks.

4.4.1. Paired-comparison task

Following Bradlow et al. (1997), a group of English NLs performed a paired-comparison task with the learners' pretest and post-test productions. This task was designed to determine if NLs judged the post-test productions of the experimental group as more native-like than pretest productions, but not those of the control group. In each trial, the target word was presented on a screen for 500 ms after which NLs heard a pretest production and a post-test production of the learner, separated by 500 ms of silence. NLs used a 7-point scale to judge which of the two productions was 'better,' or, following Bradlow et al., which was a "clearer and more intelligible pronunciation of the word shown on the screen" (p. 2303). A response of '1' indicated that the first version was better than the second, a response of '4' indicated no noticeable differences between the two versions, and a response of '7' indicated that the second version was better than the first. NLs were instructed to use all seven points on the rating scale.

NLs heard both the experimental (48 minimal pairs) and filler (28 minimal pairs) stimuli. Because there were 152 words ($[64+12]\times 2$) in each context (isolated word, before a vowel, before a consonant), this resulted in a total of 456 trials for each listener. The words from each context were separated into three blocks. Each block began with five practice items to familiarize NLs with the procedure and lasted approximately 10–12 min. Words were balanced such that in half of the cases, the pretest version preceded the post-test version, and in the other half, the post-test version preceded the pretest version. Thus, for each L2 learner, two lists were created such that in one version, the pretest was presented first and in the other, the post-test was first. Each learner was assigned two NLs. Stimuli were presented via Paradigm Player. NLs completed this task in the lab wearing either Beyerdynamic DT 770 or Sony MDR 7506 headphones and had control over the volume level via an Alesis iO2 USB interface.

In the data analysis, to facilitate the interpretation of the results, scores were converted from a scale of 1–7 to a scale of −3 to 3 such that a negative score indicated a preference for the pretest item and a positive score indicated a preference for a post-test item. Average scores were calculated across learners, taking into consideration that some learners were missing data points (see Footnote 3). To maintain consistency with the perception results, below we report only the results for words extracted from sentence contexts.

4.4.2. Forced-choice word-identification task

Following Bradlow et al. (1997), learners' productions were also presented to NLs in a forced-choice word-identification task. While the paired-comparison task can tell us whether the experimental group's post-test productions were judged as more native-like than their pretest productions, it does not provide us information about whether NLs can more accurately identify learners' post-test productions as the target word. NLs thus completed a forced-choice word-identification task similar to the one learners completed in the perception task. At the beginning of each trial, NLs saw the two words from each pair presented on the left and right side of the screen for 500 ms. Then, a version of the word was played and NLs were asked to choose the correct response as quickly as possible by pressing one of two marked keys on the keyboard.

NLs heard both the experimental (48 minimal pairs) and filler (28 minimal pairs) stimuli. The words from each context were separated into three blocks such that listeners completed three tasks. Each task began with six practice items to familiarize NLs with the procedure. Stimuli were presented using E-Prime, and participants wore either Beyerdynamic DT 770 or Sony MDR 7506 headphones and had control over the volume level via an Alesis iO2 USB interface. Each task lasted approximately 10 min. Tasks were balanced such half the stimuli were pretest versions and half were post-test versions. Thus, for each L2 learner, two lists were created. Each L2 learner was assigned two NLs. As with the perception and paired-comparison results, we report only the results for the sentential context.

5. Results

Results are presented in the sections below, beginning with the perception tasks, followed by the production task and ending with a comparison of perception and production. As mentioned earlier, the data reported on in the current study pertain only to the sentence contexts.

5.1. Perception results

Five native English listeners completed the perception pretest. The average d' perception score for real words was 4.19 (SD = 0.49) and that for nonce words was 4.42 (SD = 0.33). A paired-samples t-test revealed no significant difference between real and nonce words for the native English listeners, $t(4) = -1.58$, $p < 0.190$. The task was therefore adequate to test for the perception of palatal codas.

Fig. 1 presents pretest and post-test d' perception scores of the experimental and control groups separately for real and nonce words. For the experimental group, these scores conflate familiar and new words (i.e., words that had vs. had not been heard in the training) and familiar and new talkers (i.e., talkers that had vs. had not been heard in the training), whereas for the control group, all words and all talkers were new insofar as their training did not include words or talkers from the pre- and post-test.

To determine whether experimental and control groups performed differently on the real and nonce words in the pretest and post-test, a mixed-design repeated-measures ANOVA was performed with test (pretest, post-test) and word type (real, nonce) as within-subject variables and with group (experimental, control) as between-subject variable. It revealed a main effect of test, $F(1,22) = 22.53$; $p < 0.001$, but no main effect of word type, $F < 1$. There was also an interaction between test and group, $F(1,22) = 9.39$; $p < 0.006$, but no interaction between word type and group, $F < 1$, or main effect of group, $F(1,22) = 2.51$; $p < 0.128$. Paired-samples t-tests were thus conducted on each group's pretest and post-test d' scores collapsed across real and nonce words (since word type did not interact with group), with alpha levels adjusted to $p < 0.025$. They revealed a significant difference between the pretest and post-test scores for

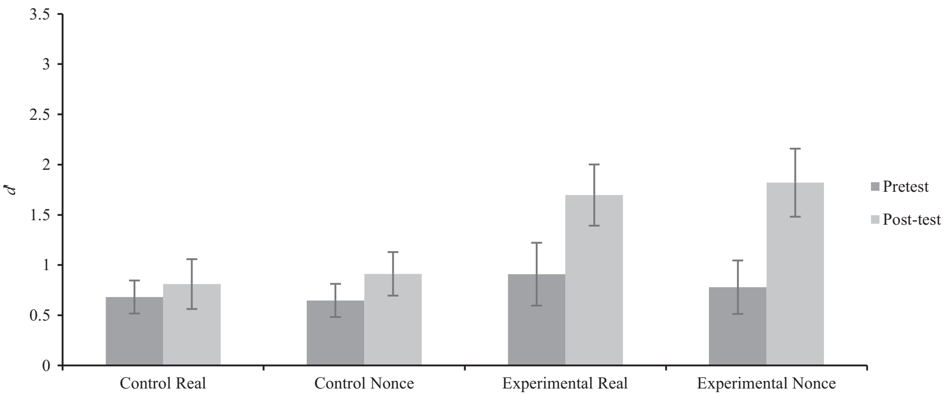

Fig. 1. Pretest and post-test mean d' perception scores (and standard error bars) by group and word type.

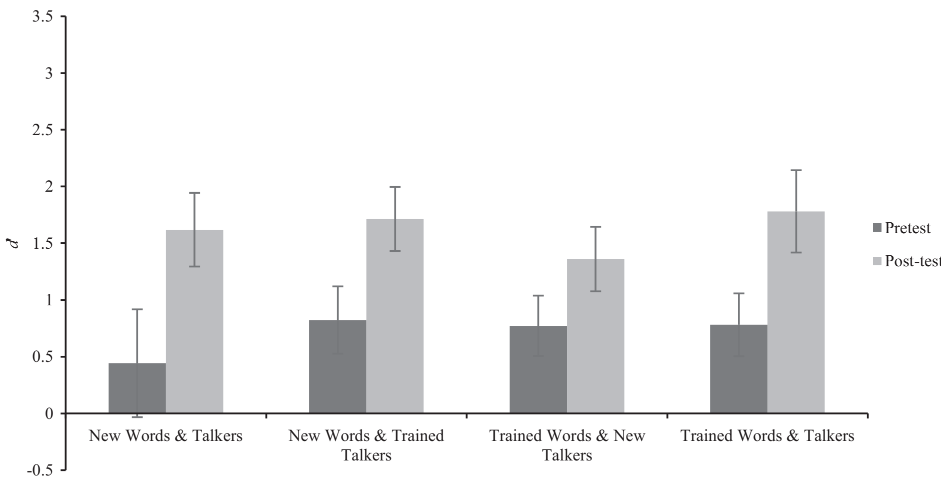

Fig. 2. Pretest and post-test mean d' perception scores (and standard error bars) for the experimental group with new and trained words and talkers.

the experimental group, $t(11) = -4.96$; $p < 0.001$, but not for the control group, $t(11) = -1.50$; $p < 0.161$. Thus, the experimental group showed significant improvement between the pretest and post-test for palatal codas in sentences, but the control group did not.

To ensure that the overall lack of improvement of the control group was not a result of the smaller amount of time they spent on training, for that group we examined the relationship between the individual participants' perception improvement scores (i.e., the differences between the post-test and pretest scores) and the amount of time they spent on perceptual training (i.e., from 28 to 127 min). The correlation analysis did not reach significance and in fact was in the opposite direction to what would be predicted if the control training had resulted in participants' improved performance ($r = -0.304$; $p < 0.337$). Thus, it is unlikely to be the case that participants in the control group performed less accurately than those the experimental group because they spent less time on perceptual training. These results suggest that perceptual training on palatal codas, and not time spent on task, resulted in the significant improvements demonstrated by the experimental group.

Results from the experimental group were then analyzed separately to determine whether participants were able to generalize learning to new words and new talkers. We compared the experimental group's pretest and post-test d' scores on four categories of words: (1) new words spoken by new talkers, (2) new words spoken by talkers from the training, (3) words from the training spoken by new talkers, and (4) words from the training spoken by talkers from the training. Fig. 2 shows the experimental group's pretest and post-test d' scores for words produced by new and trained words and talkers.

A repeated-measures ANOVA performed with test (pretest, post-test), word (new, trained), and talker (new, trained) as within-subject variables revealed a main effect of test, $F(1,11) = 27.43$; $p < 0.001$, but no other main effects or interactions (talker: $F(1,11) = 2.99$; $p < 0.112$; word × talker: $F(1,11) = 4.09$; $p < 0.068$; all other F's < 1). Thus, learners showed similar improvements between pretests and post-tests on all four categories of words. From these results, we can conclude that participants were able to generalize learning to new words and new talkers.

In summary, the experimental group, but not the control group, showed significant perception improvement between the pretest and post-test in the sentence context. These findings suggest a beneficial effect of perceptual phonetic training on palatal codas. Results of generalizability indicated that the experimental group was able to generalize learning to new words and new talkers, suggesting that the perceptual training was successful in allowing learners to establish more robust representations. Finally, no effect of lexical status was found for the perception results of either group, indicating that word familiarity did not influence the results. We now turn our attention to the production results.

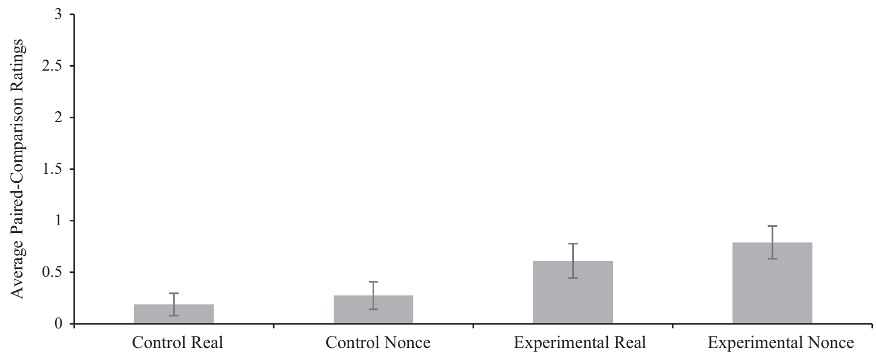

Fig. 3. Mean paired-comparison ratings (and standard error bars) by group and word type.

5.2. Production results

5.2.1. Paired-comparison task

For this task, NLs were presented with both the pretest and post-test target word productions of a learner and judged which was more native-like; thus, as NLs chose only one word, results cannot be separated by pretest and post-test. Fig. 3 displays the paired-comparison production results.

A mixed-design repeated-measures ANOVA with word type (real, nonce) as within-subject variable and group (experimental, control) as between-subject variable revealed a main effect of word type, $F(1,22)=9.91$; $p<0.005$, and a main effect of group, $F(1,22)=5.60$; $p<0.027$, but no interaction between word type and group, $F(1,22)=1.21$; $p<0.283$. Thus, more of the experimental group's post-test productions of palatal codas in the sentence context were rated more native-like in comparison to those of the control group.

These results indicate that the perceptual phonetic training on palatal codas helped learners improve their production of palatals, and improvements, as measured by the paired-comparison task, were minimal for the control group. These results also show that improvements were greater for nonce words than real words. One possible explanation for this finding is that the NLs may have been more lenient in their judgments of nonce words as opposed to real words because for the former there was no top down information to influence their decisions. Word familiarity has been shown to influence the degree to which listeners attend to acoustic cues or allow top-down information to influence perception in L2 studies (e.g., Flege, Takagi, & Mann, 1996; MacKay, 1987).

Another aspect of the paired-comparison results to consider is that averages for the control group are above zero, indicating a slight preference for post-test productions. This is not surprising, however, if we consider that the control group also received perceptual phonetic training on vowels. This finding does not imply that NLs were able to better identify the control group's post-test productions, but rather that they indicated some of their post-test productions were more native-like than their pretest productions. The phonetic training the control group received on vowels could have made their post-test productions more native-like. Nevertheless, the significant difference between the experimental and control group demonstrates that perceptual phonetic training on palatal codas contributed to significantly more native-like judgments from NLs for the experimental group. It should also be noted, however, that the positive values for the experimental group (0.6 and 0.8 for real and nonce words, respectively), are still relatively small. Thus, while there was production improvement by perceptual training, it was limited.

We now turn our attention to the results of the word-identification task to determine whether NLs were able to categorize the experimental group's post-test productions more accurately.

5.2.2. Forced-choice word-identification task

Fig. 4 displays the d' scores of NL's ratings of pretest and post-test productions of the experimental and control groups separately for real and nonce words.

A mixed-design repeated-measures ANOVA was performed with test (pretest, post-test) and word type (real, nonce) as within-subject variables and with group (experimental, control) as between-subject variable. There was a main effect of test, $F(1,22)=17.22$; $p<0.001$, and an interaction between test and group, $F(1,22)=7.05$; $p<0.014$, but no main effect of word type, $F<1$, or group, $F(1,22)=3.58$; $p<0.072$, and no interaction between word type and group, $F(1,22)=1.16$; $p<0.293$, or between word type and test or word type, test, and group, $F<1$. To further examine the effect of test, paired-samples t-tests were conducted on the d' scores collapsed across real and nonce words (since word type did not interact with group), with alpha levels adjusted to $p<0.025$. They revealed a significant difference between the pretest and post-test scores for the experimental group, $t(11)=-4.10$; $p<0.002$, but not for the control group, $t(11)=-1.41$; $p<0.186$. Thus, for the experimental group, NLs were able to accurately identify more post-test productions than pretest productions. The same was not true for the control group.

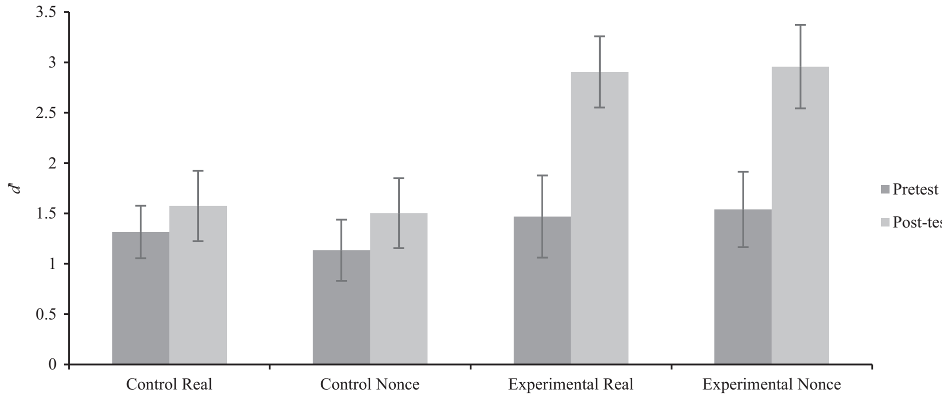

Fig. 4. Pretest and post-test mean *d′* production scores (and standard error bars) from the word-identification task by group and word type.

Table 2
Pretest, post-test, and improvement scores (as *d′* scores) by participant.

Participant	Palatal Codas: Sentences					
	Perception			Production		
	Pretest	Post-test	Improvement	Pretest	Post-test	Improvement
P1	0.21	0.84	0.63	0.60	3.53	2.93
P2	0.16	2.11	1.95	2.45	3.41	0.96
P3	0.24	1.53	1.29	0.15	3.37	3.22
P4	0.37	1.66	1.30	1.58	3.27	1.70
P5	0.48	1.19	0.71	0.66	4.06	3.41
P6	0.06	0.06	0.01	0.06	0.59	0.53
P7	0.72	2.13	1.41	1.87	2.79	0.92
P8	−0.22	0.00	0.22	−1.30[a]	0.07	1.37
P9	1.38	3.06	1.68	3.27	4.09	0.82
P10	2.72	2.66	−0.06	2.54	2.80	0.25
P11	1.29	2.12	0.83	2.06	3.78	1.72
P12	2.33	3.48	1.15	3.37	2.86	−0.51
Mean	**0.81**	**1.74**	**0.93**	**1.44**	**2.89**	**1.44**

[a] The large negative *d′* score of Participant 8's perception pretest can be explained by the fact that this participant *almost* always produced CVCy and CVC sequences as CVC. If it were the case that the participant *only* produced CVC sequences, the *d′* would be 0 (no discrimination). However, because this participant showed some variation (specifically, she produced one CVC sequence as CVCy and one CVCy sequence as CVCy), this resulted in the large negative value. While large negative *d′* values might be explained by a participant purposefully giving the wrong answer (cf. Park, 2013), we do not believe this is the case with this particular participant.

In summary, phonetic training on palatal codas enhanced the learners' production of codas, as reflected in NLs' identification of the produced words. However, unlike the paired-comparison task, there was no difference between real and nonce words in the forced-choice word-identification task, suggesting that the effect of word type in the paired-comparison task may indeed have been due to NLs' greater leniency in judging the nonce words produced by the learners. We now turn to an examination of individual variability and the relationship between perception and production gains.

5.3. *Individual variability and the relationship between perception and production*

In this subsection, we consider the relationship between the perception and production changes that phonetic training on palatal codas incurred for the experimental group. Assuming that NLs' identification of the words produced by the experimental group provided a more accurate representation of the learners' accuracy in producing palatal codas, we use this measure as estimate of the learners' production improvement between the pretest and post-test. Table 2 shows the pretest, post-test, and improvement scores (as *d′* scores) on the perception and (word-identification) production tasks for each participant. Improvement scores were calculated by subtracting pretest scores from post-test scores: d' (improvement) = d' (post) − d' (pre).

As can be seen in Table 2, all but one participant (P10) showed improvement in the perception of palatal codas in sentences. This participant happened to have the highest pretest score among learners in the group. A correlation analysis between the perception pretest scores and perception improvement scores did not reach significance ($r=-0.087$; $p<0.788$), suggesting that the learners' amount of perception gain may not be related to their accuracy on the pretest. Similarly, when we consider production scores, we see that all but one learner (P12) improved on their production of palatal codas. This participant's pretest score was also the highest among the group. A correlation analysis between the production pretest scores and production improvement scores also did not reach significance ($r=-0.542$; $p<0.068$).

To shed light on the relationship between perception and production, a correlation analysis between the improvement scores on the perception task and those on the production task (word-identification measure) was conducted. This analysis did not yield a significant relationship between the two ($r=0.029$; $p<0.929$). Thus, despite finding that perceptual training contributed to enhancing both the perception and production of palatal codas in Korean L2 learners of English, we do not see a direct relationship between the two. We now turn to a discussion of these findings and their implications for understanding the nature of the representations underlying the L2 perception of syllable structure.

6. Discussion

This study set out to investigate the mediating effect of perceptual training on the perception and production of syllable structure, thereby shedding light on the nature of stored representations and on the relationship between L2 speech perception and production. The first important finding is that the beneficial effects of perceptual training can extend to syllable structure. The results indicated that the experimental group, but not the control group, improved their perception of palatal codas in English. Furthermore, the experimental group was able to generalize improvements to both new words and new talkers, and improvements in perception were found for both real and nonce words. This is in line with previous work showing that perceptual phonetic training is beneficial for establishing new segmental categories (e.g., Bradlow et al., 1997; Lambacher et al., 1995; Lively et al., 1993; Logan et al., 1991). Taken together, these findings indicate that the perceptual phonetic training used in this study can yield robust improvements in learners' perception and production of palatal codas, with this improvement not being word- or talker-specific.

The second important finding of this study is that perceptual training on palatal codas enhanced learners' production of palatal codas. These results are consistent with those of previous research showing that perceptual training on difficult segments can yield production improvements for the same segments (e.g., Bradlow et al., 1997; Lively et al., 1993). This suggests that perception and production systems must be connected in some way. Despite this mediating effect, however, the results showed no correlation between learners' improvements in the perception and production of palatal codas. This lack of relationship cannot be attributed to baseline differences in L2 learners' perception and production accuracies, as no relationship was found between their overall accuracy in the pretests and the amount of improvement they made in the post-tests. The absence of a direct link between perception and production improvements suggests that perception and production do not bear a one-to-one relationship. This finding is also in line with much of the previous research that examined the relationship between perception and production (Bohn & Flege, 1990; Bradlow et al., 1997; De Jong, Hao, & Park, 2009; Flege & Eefting, 1987; Goto, 1971; Sheldon & Strange, 1982).

Our results can be accounted for straightforwardly if we assume that (at least some of) the representations underlying learners' speech perception and speech production differ in kind. As discussed earlier, a model couched within a Direct Realist theory of speech perception (e.g., PAM(-L2); Best, 1995; Best & Tyler, 2007) would posit linked perception and production systems that share representations, with these representations being gestural in nature; thus, such a model would predict that perception and production learning would be directly related. By contrast, a model situated within a psychoacoustic theory of speech perception (e.g., the SLM; Flege, 1995) would propose that at least some of the representations underlying speech perception and speech production are different, with the former being psychoacoustic and the latter being articulatory; according to such a model, perception and production learning would not be directly related. Given that our learners' perception and production of syllable structures do not develop hand in hand, our results are more easily explained by the second type of model, with perceptual representations being psychoacoustic in nature and thus not explaining all the variability in the production results.

Bradlow et al. (1997) explained asymmetries between improvements in perception and improvements in production in their results in several ways. They considered the possibility that some learners' production improvements did not match those of participants with similar perception improvements because they needed more time to acquire the motor skills to produce the /ɹ/-/l/ contrast. They also considered the possibility that some learners were attending to cues during training that would not aid in improving perception but that might aid in improving production. An important body of research indeed suggests that L2 learners' speech perception difficulties stem in large part from their non-target-like weighting of acoustic cues to phonetic categories (e.g., Holt & Lotto, 2006; Idemaru, Holt, & Seltman, 2012; Iverson et al., 2005; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; Lotto, Sato, & Diehl, 2004; Mirman, Holt, & McClelland, 2004). In Bradlow et al. (1997), the most important acoustic cue for perceiving the relevant phonetic contrast (/ɹ/-/l/) was a place of articulation cue (F3). The authors argued that some learners may have focused on durational rather than spectral cues to produce the contrast, but these durational cues may not have been sufficient to perceive the contrast.

Our study differs from Bradlow et al.'s (1997) in that the contrast pairs differed more in the timing of the articulatory gestures than in their place of articulation. In order to develop native-like perception of palatal codas, learners must attend to the cues that signal these timing differences. These cues include the duration of the stem vowel and the duration of the syllable coda/onset consonant. The perceptual improvements we observed in almost all learners in the experimental group suggest that they indeed attended to these cues. At the same time, our individual L2 learners improved perception and production at different rates. Some learners may thus have tuned more to the perceptual cues to the realization of palatal codas in English. By contrast, others may have tuned more to the mapping between these cues and the timing of articulatory gestures in speech production. It is precisely at the level of the mapping between acoustic cues to syllable structure and the timing of articulatory gestures in the production of syllable structure that perceptual training may have enhanced the learners' production of codas: By hearing a large number of phonetically variable English words ending with a palatal coda, learners in the experimental group were better able to approximate the acoustic targets for such codas in their speech production. The learners who improved on their perception but not on their production of palatal codas did not appear to have tuned to this mapping and focused their attention strictly on the perceptual correlates of syllable structure information. Taken together, these results provide at least some support for the view that the timing of articulatory commands may be somewhat independent from the creation of perceptual representations.

Although this study did not investigate the status of learners' intrusive vowels, our explanation of the results is compatible with Davidson's (2006, 2010) account, according to which non-native speakers' production of intrusive vowels are the result of mistiming of articulatory gestures. If, in the present study, illusory vowels in speech perception and intrusive vowels in speech production both resulted from L2 learners' repair of syllable structure due to L1 syllable-structure restrictions, we might also have expected perceptual and production improvements to develop together, contrary to fact. It is unclear whether illusory vowels were heard in our word-identification task because learners attempted to repair the syllable structure they heard due to L1 restrictions or because they were influenced by L1-specific phonetic cues. Either way, the differential development of speech perception and speech production at the individual level suggests that at least in production, it is unlikely that Korean L2 learners of English insert intrusive vowels strictly for phonological reasons.

This study leaves a number of questions open for further investigation. One is how the different perceptual cues to palatal codas in English influence Korean listeners' perception of these codas. The present study did not focus on these individual cues because it did not aim to tease apart an L1 syllable-structure restriction account from an L1-specific phonetic-cue account. Further research should examine the nature of the L1 influence observed in this and other studies on the perception of syllables structure. Another question is whether the cues that enhance learners' perception of syllable structure after perceptual training are also produced after such a training. This is something that future research should also examine, as it can shed further light on the nature of the relationship between L2 speech perception and L2 speech production.

Another question that remains open for further research is whether similar results would be obtained with participants at lower proficiencies. Our Korean L2 learners of English had proficiency levels that ranged from intermediate to advanced; they had studied English for many years and lived in an immersion context for an extended period of time. It is therefore unclear how these learners would perceive palatal codas in English at the outset of acquisition. Future work investigating the performance of lower-level learners might yield different results from those reported here. For example, one might find for low-level learners, perceptual training results in much more perceptual learning than in production learning. This would suggests that perceptual representations develop more rapidly than the corresponding articulatory commands early on in development. Research focusing on lower-level learners can thus provide important insights for speech perception and speech production models.

7. Conclusion

The current research contributes to a better understanding of the relationship between perception and production systems. It extends previous work by adding a perceptual training beyond isolated words, and demonstrates improvements also occur in these contexts. It also provides evidence that perceptual phonetic training can be beneficial not only for acquiring new segment contrasts, but also for acquiring segments in restricted syllable structures. Finally, it adds to a growing body of literature which provides evidence that representations are not directly shared between perception and production systems.

Acknowledgments

We would like to thank Chris Carignan for his help in developing and modifying the scripts used to automate processes for segmenting and normalizing data in Praat. We would also like to thank Liam Moran, a consultant for ATLAS Digital Media at the University of Illinois at Urbana-Champaign, who created the Pierceive software and Lisa Pierce at the University of Illinois at Urbana-Champaign, who allowed us access to the training paradigm she created for the purposes of using it as the control group training. We would also like to thank Editor Dr. Erika Levy as well as two anonymous reviewers for their valuable comments on our manuscript.

Appendix. Experiment stimuli

/ʤ/
1. edge/edgy
2. dodge/dodgy
3. smudge/smudgy
4. hedge/hedgy
5. cage/cagey
6. fudge/fudgy
7. sludge/sludgy
8. wedge/wedgy
9. tudge/tudgy
10. pidge/pidgy
11. codge/codgy
12. bedge/bedgy
13. modge/modgy
14. leidge/leidgy
15. sodge/sodgy
16. feidge/feidgy

/ʃ/
17. push/pushy
18. ash/ashy
19. trash/trashy
20. fish/fishy
21. bush/bushy
22. flash/flashy
23. slush/slushy
24. mush/mushy
25. teesh/teeshy
26. pash/pashy
27. cosh/coshy
28. bosh/boshy
29. mish/mishy
30. leish/leishy
31. seish/seishy
32. fush/fushy

/ʧ/
33. catch/catchy
34. itch/itchy
35. sketch/sketchy
36. stretch/stretchy
37. peach/peachy
38. twitch/twitchy
39. patch/patchy
40. touch/touchy
41. tetch/tetchy
42. putch/putchy
43. petch/petchy
44. boatch/boatchy
45. mutch/mutchy
46. letch/letchy
47. sotch/sotchy
48. fatch/fatchy

Fillers
49. fend/pend
50. chief/cheap
51. flake/fleck
52. cologne/clone
53. train/terrain
54. drive/derive
55. polite/plight
56. filet/flay
57. parade/prayed
58. beret/bray
59. sale/sell
60. miss/mist
61. pass/past
62. blow/below
63. blike/belike
64. pleam/paleam
65. fape/pape
66. heff/hepp
67. tabe/tebb
68. clate/calate
69. treem/tereem
70. prume/perume
71. prace/perace
72. mape/mepp
73. tiss/tissed
74. rass/rassed
75. froy/feroy
76. drate/derate

References

Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–204). Timonium, MD: York Press.

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

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775–794.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591–630.

Bohn, O. S., & Flege, J. E. (1990). Perception and production of a new vowel category by adult second language learners. In J. Leather, & A. James (Eds.), *Proceedings of the 1990 Amsterdam symposium on the acquisition of second-language speech* (pp. 37–56). Amsterdam: University of Amsterdam.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.2.04) [Computer program]. Retrieved from 〈http://www.praat.org/〉.

Bradlow, A., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.

Brannen, K. (2002). The role of perception in differential substitution. *Canadian Journal of Linguistics*, *47*, 1–46.

Broselow, E. (2009). Stress adaptation in loanword phonology: Perception and learnability. In P. Boersma, & S. Hamann (Eds.), *Phonology in perception* (pp. 191–234). Berlin: Mouton de Gruyter.

Broselow, E., Chen, S. I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261–280.

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research*, *7*, 35–59.

Carlisle, R. S. (1997). The modification of onsets in a markedness relationship: Testing the interlanguage structural conformity hypothesis. *Language Learning*, *47*, 327–361.

Carlisle, R. S. (1998). The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition*, *20*, 245–260.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*, 489–534.

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*, *19*, 619–633.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*, 104–137.
Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, *38*, 272–288.
Davidson, L. (2011). Phonetic, phonemic and phonological factors in cross-language discrimination of phonotactic constraints. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 270–282.
Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, *40*, 234–248.
Davidson, L., & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. In G. Garding, & M. Tsujimura (Eds.), *WCCFL 22 proceedings* (pp. 165–178). Somerville, MA: Cascadilla Press.
Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635–647.
De Jong, K., Hao, Y.-C., & 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.
Diehl, R. L. (1987). On segments and segment boundaries. *Journal of Phonetics*, *15*, 289–290.
Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*, 121–144.
Dupoux, E., Kakehi, K., Hirose, Y., Pallier, P., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion?. *Journal of Experimental Psychology*, *25*, 1568–1578.
Dupoux, E., Pallier, C., Kaheki, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, *16*, 491–505.
Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from?. *Journal of Memory and Language*, *64*, 199–210.
Eckman, F., & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research*, *9*, 234–252.
Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, *121*, 3814–3826.
Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner, & C. W. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249–289). Amsterdam: John Benjamins.
Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). Timonium, MD: York Press.
Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–335). Berlin: Mouton de Gruyter.
Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.
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.
Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973–2987.
Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, *38*, 25–55.
Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161–1173.
Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.
Fowler, C. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology*, *1*, 145–160.
Fowler, C. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America*, *89*, 2910–2915.
Fowler, C. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception and Psychophysics*, *55*, 597–610.
Fowler, C. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1730–1741.
Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*, *9*, 317–323.
Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation?. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 592–608.
Hancin-Bhatt, B. (2000). Optimality in second language phonology: Codas in Thai ESL. *Second Language Research*, *16*, 201–232.
Hancin-Bhatt, B., & Bhatt, R. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second language Acquisition*, *19*, 331–378.
Holt, L. L., Lotto, A. J., & Kluender, K. R. (1998). Incorporating principles of general learning in theories of language acquisition. In M. Gruber, C. Derrick, K. S. Higgins, Olson, & T. Wysocki (Eds.), *Chicago Linguistic Society, Volume 34: The Panels* (pp. 253–268). Chicago, IL: Chicago Linguistic Society.
Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*, 3059–3071.
Idemaru, K., Holt, L. L., & Seltman, H. (2012). Relative duration or absolute duration? Cue weighting in the perception and production of Japanese stops length distinction. *Journal of the Acoustical Society of America*, *132*, 3950–3964.
Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2011). Can native Japanese listeners learn to differentiate /r-l/ on the basis of *F*3 onset frequency?. *Bilingualism: Language and Cognition*, http://dx.doi.org/10.1017/S1366728911000447.
Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, *39*, 571–584.
Iverson, P., & Evans, B. G. (2007). Auditory training of English vowels for first-language speakers of Spanish and German. In *Proceedings of the 16th international congress of phonetic sciences*. Saarbrücken, Germany (pp. 1625–1628).
Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching /r/-/l/. *Journal of the Acoustical Society of America*, *118*, 3267–3278.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.
Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints?. *Language and Speech*, *50*, 23–52.
Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.
Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173–217). San Diego, CA: Academic.
Kim, K. (2009). Coronals in epenthesis in loanwords with comparisons to velars. *Toronto Working Papers in Linguistics*, *30*, 53–67.
Kuhl, P. K. (1985). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology*, *45*, 233–265.
Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*, 227–247.
Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify /l/ and /r/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.
Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.
Lotto, A. J. (2000). Language acquisition as complex category formation. *Phonetica*, *57*, 189–196.
Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. C181–C186). Cambridge, MA: MIT Research Laboratory in Electronics.
MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *TRENDS in Cognitive Science*, *11*, 442–450.
McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the /r/–/l/ contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience*, *2*, 89–108.
Mirman, D., Holt, L. L., & McClelland, J. M. (2004). Categorization and discrimination of non-speech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America*, *116*, 1198–1207.
Neary, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, *18*, 347–373.
Oh, E. (2008). Coarticulation in non-native speakers of English and French: An acoustic study. *Journal of Phonetics*, *36*, 361–384.
Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1718–1725.
Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, *41*, 78–87.
Peperkamp, S., & Dupoux, E. (2003). Reinterpreting loanword adaptations: The role of perception. In *Proceedings of the 15th international congress of phonetics science* (pp. 367–370).
Perception Research Systems. (2007). *Paradigm* [Computer Program]. Retrieved from ⟨http://www.paradigmexperiments.com⟩.
Rochet, B. L. (1995). Perception and production of second language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–204). Timonium, MD: York Press.
Schmidt, A., & Meyer, K. (1995). Traditional and phonological treatment for teaching English fricatives and affricates to Koreans. *Journal of Speech and Hearing Research*, *38*, 828–838.

Schneider, W., Eschman, A., & Zuccolotto, A. (2001a). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.
Schneider, W., Eschman, A., & Zuccolotto, A. (2001b). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.
Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*, 243–261.
Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*, 1272–1278.
Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: 'Clozing' the gap. *Studies in second language acquisition*, *33*, 339–372.
Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, *44*, 121–133.
Yeon, S. H. (2004). *Teaching English word-final alveopalatals to native speakers of Korean* (Doctoral dissertation). University of Florida.
Zsiga, E. (2003). Articulatory timing in a second language. *Studies in Second Language Acquisition*, *25*, 399–432.