LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Effects of Different Types of Corrective Feedback on Receptive Skills in a Second Language: A Speech Perception Training Study

Andrew H. Lee and Roy Lyster
McGill University

This study investigated the effects of different types of corrective feedback (CF) provided during second language (L2) speech perception training. One hundred Korean learners of L2 English, randomly assigned to five groups (n = 20 per group), participated in eight computer-assisted perception training sessions targeting two minimal pairs of English vowels. Four treatment groups each received a different type of CF; three groups received one of three types of auditory CF and a fourth group received a visual type of CF; the control group did not receive CF. Results of pretests, immediate posttests, and delayed posttests showed that, in comparison to the control group, the groups that received auditory CF improved significantly in trained over untrained words, whereas

This research was supported by a Social Sciences and Humanities Research Council of Canada grant (410-2011-0671) awarded to Roy Lyster. Earlier versions of this article were presented at the joint meeting of the American Association for Applied Linguistics and the Canadian Association of Applied Linguistics held in Toronto, ON, in March 2015, and at the Second Language Research Forum held in Atlanta, GA, in October 2015. We sincerely thank all participants and extend our gratitude to the anonymous reviewers as well as Editor Pavel Trofimovich and Associate Editor Emma Marsden for providing useful feedback on earlier drafts. We are also grateful to the following research assistants, who contributed at various phases of this study: Benjamin Gormley, James Mutch, and Sunyoung Park.

This article has been awarded an Open Materials badge. All materials are publicly accessible in the IRIS digital repository at http://www.iris-database.org. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Andrew H. Lee, Department of Integrated Studies in Education, McGill University, 3700 McTavish Street, Montreal, QC H3A 1Y2, Canada. E-mail: andrew.lee@mcgill.ca

DOI: 10.1111/lang.12167

the group that received visual CF fared less well. These results are discussed in terms of the benefits of auditory CF types, especially CF combining target and nontarget forms.

Keywords corrective feedback; second language; speech perception training; speech learning

Introduction

Many descriptive and experimental studies have investigated corrective feedback (CF) in relation to second language (L2) learners' oral production, while others have been concerned with CF on written production. In both cases, the focus of the research has been mainly on learner productive skills. Building upon and moving beyond studies targeting CF on learner productive skills, the present study aimed to examine different types of CF and their effects in one domain of receptive skills, namely, L2 speech perception.

The effectiveness of CF on oral production has been demonstrated through both meta-analyses (Li, 2010; Lyster & Saito, 2010) and narrative reviews (Lyster, Saito, & Sato, 2013). The effects of CF on learner production have been investigated primarily with respect to morphosyntactic targets, and more recent studies have targeted the effects of CF on phonological development. For instance, Saito and Lyster (2012a, 2012b) and Saito (2013a) found that Japanese learners of L2 English benefited from recasts. Saito and Wu (2014) also found positive effects for CF on L2 tonal acquisition. In the same vein, Dlaska and Krekeler (2013) concluded that individualized CF is a significant and powerful teaching tool in L2 pronunciation instruction.

Concerning speech perception, a number of L2 speech perception training studies (Hardison, 2003; Logan, Lively, & Pisoni, 1991; McClelland, Fiez, & McCandliss, 2002; Wang & Munro, 2004) have emphasized the benefits of CF in laboratory-based training sessions, and Lee and Lyster (2015) demonstrated the effects of CF in classroom-based L2 speech perception training. However, these studies adopted one or two types of CF and did not compare the effects of different types of CF. In particular, the type of CF in laboratory-based L2 speech perception training has been uniform and simple, conveying a right-or-wrong message (Derwing & Munro, 2015). Given that different types of CF trigger different types of cognitive processes (e.g., Lyster & Izquierdo, 2009), such a methodological limitation begs the question as to whether various types of CF could be applicable to L2 speech perception training and whether they would yield differential effects on development. Consequently, the present study explored the acquisitional value of different types of CF on L2 learners' receptive skills, with the goal of providing insight into the role of CF in L2 speech perception training.

Background

Speech Perception in L2 Phonological Learning

Many scholars have argued that the ultimate goal of L2 pronunciation learning should be intelligible rather than nativelike speech (Derwing & Munro, 2005, 2009; Field, 2005; Levis, 2005; Setter & Jenkins, 2005). In terms of achieving intelligible speech, one of the key factors discussed in L2 phonology concerns the perceptual difficulties that learners commonly encounter (Escudero, 2006; Flege, 1995, 2003). As Borden, Gerber, and Milsark (1983) and Llisterri (1995) contended, accurate perception is a requisite ability for optimal L2 speech production.

The importance of perception in speech acquisition has been supported by several theoretical and empirical accounts. As for theoretical accounts, Best's (Best, 1995; Best & Tyler, 2007) perceptual assimilation model (PAM), which was derived from the frameworks of articulatory phonology (Browman & Goldstein, 1989) and direct realism (Fowler, 1986), predicts the initial state of L2 speech learning, in terms of learners' perception of nonnative phonemes (sounds, for short). According to PAM, it is linguistic experience that leads infants to successful speech perception in their native language (L1). When listeners perceive a particular sound, they recognize and extract relevant articulatory (gestural) patterns from the speech signal; in this way, infants who are learning their L1 build a speech perception system specific to the input they receive. However, with regard to L2 learners, their L1 perception system often makes it hard for them to detect L2-specific articulatory gestures for nonnative sounds, such that nonnative sounds are assimilated to, or perceived in terms of, existing L1 sound categories. Consequently, depending on specific assimilation patterns, learners can be good at categorizing L2 sounds, for example, when each of two contrasting nonnative sounds is associated with a different L1 sound (two-category assimilation), or learners can have considerable perceptual difficulties, for instance, when two contrasting sounds are subsumed by one L1 category (single-category assimilation). PAM posits that speech production and perception are closely related, in that both modalities share representations, processes, and resources. Listeners exploit articulatory gestures as the basis of speech perception.

On the other hand, Flege's (1995) speech learning model (SLM) explicitly attributes successful L2 production to accurate L2 perception, suggesting that perception precedes production. One of the postulates of this model is that "[t]he mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning" (p. 239). However, according to SLM, application of learning mechanisms to L2 learning is constrained by several factors, one of which is perceived cross-linguistic similarity. When L2 learners experience nonnative sounds, learners classify them as new or similar. For example, nonnative sounds that do not exist in learners' L1 sound systems are defined as new, while similar sounds are those that have perceptually similar equivalents in L1 sound systems. In this regard, L2 learners are predicted to have more difficulty perceiving similar sounds than new sounds, and they need more time and intervention to acquire similar sounds compared to new sounds.

With respect to empirical accounts, a vast number of experimental studies have advocated the importance of perception in L2 speech acquisition, most of which investigated the influence of perception training on production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Jamieson & Rvachew, 1992; Rochet, 1995; Rvachew & Jamieson, 1995; Wang, Jongman, & Sereno, 2003). For instance, Bradlow et al. (1997) found that Japanese learners of L2 English improved their production ability regarding the English /ɹ/-/l/ contrast only after perception training. Similarly, Hardison (2003) revealed positive transfer of perception training to production performance with Korean and Japanese learners of L2 English. In addition, neurolinguistic findings have shown that speech perception involves speech production. For example, Watkins and Paus (2004) found that there is robust excitability of the motor system controlling speech production during auditory speech perception. Furthermore, Pulvermüller and Schumann (1994) demonstrated that neurons controlling the movement of relevant muscles during sound articulation are also activated with the sensory neurons in the auditory cortex as it processes the sound via the auditory system. In light of these theoretical and empirical accounts, speech perception can be considered to play a key role in L2 speech learning, especially with respect to the influence of perception on production. If this is the case, what are effective training techniques for L2 speech perception?

L2 Speech Perception Training

Most L2 perception training studies, which have been conducted in laboratory settings with computer-assisted instructional interventions, have revealed positive effects on L2 learners' perceptual accuracy. The training, which tends to vary across studies, has been operationalized by means of several perception tasks, such as discrimination and identification (for review, see Strange

& Shafer, 2008). For instance, Strange and Dittmann (1984) trained Japanese learners of L2 English with synthesized and natural speech tokens. Lively, Logan, and Pisoni (1993) explored effects of multiple- versus single-talker training. Similarly, Hardison (2003) exposed L2 learners to multiple talkers through auditory-visual training. Iverson, Hazan, and Bannister (2005) provided L2 learners with different types of stimuli that had been acoustically manipulated for learners to focus on the acoustic properties of target sounds (see also Uther, Knoll, & Burnham, 2007). Training with multiple exemplars, which is referred to as high-variability perception training, is generally known to be effective in terms of improving perceptual accuracy because it pushes learners to detect crucial phonetic cues associated with target sounds as well as cues irrelevant to their category identity. In addition, training L2 learners with high-variability stimuli is believed to help learners to transfer the learned knowledge to untrained speech materials (Thomson, 2011). Moving beyond laboratory-based perception training, Lee and Lyster (2015) trained Korean learners of L2 English in a simulated classroom setting with a number of instructional techniques, such as explicit instruction, awareness tasks, and CF. Kissling (2015) also showed that L2 learners benefited from explicit phonetic instruction in terms of perceptual accuracy.

The Present Study

With respect to training methods targeting L2 perception, one important conceptual and methodological issue is whether or not training should include CF. On the one hand, language learners are known to rely on statistical learning by exploiting various regularities available in the input, for example, in order to acquire phonetic category boundaries (Baese-Berk, 2010; Saffran, Aslin, & Newport, 1996). For example, so-called implicit perceptual learning was introduced in several studies, such as those by Maye and Gerken (2000, 2001) and Baese-Berk, with no CF included in any training sessions, based on the premise that naturalistic acquisition would be achieved without CF.

However, there is little doubt that CF can serve as a catalyst for L2 development (see reviews by Lyster et al., 2013, and Sheen & Ellis, 2011). Across CF literature, there are two major types of CF: Recasts provide learners with accurate forms, whereas prompts withhold target forms, pushing leaners to modify their output. Many experimental studies have compared different types of CF (i.e., recasts vs. prompts), focusing primarily on productive skills and almost exclusively on morphosyntactic errors. As for phonological errors, a few studies investigated the effects of CF on L2 speech production, showing overall positive effects of CF on L2 output (e.g., Saito, 2013a, 2013b; Saito &

Wu, 2014). As for perception, Lee and Lyster (2015) demonstrated the effectiveness of CF provided during L2 speech perception training and concluded that CF provides learners with opportunities to retrieve, restructure, and consolidate their L2 phonological representations. In the same vein, several studies addressing the effects of perception training (e.g., Hardison, 2003, 2012; Logan et al., 1991; McClelland et al., 2002; Wang & Munro, 2004) have also highlighted the role of CF during training sessions.

Notwithstanding the importance of CF in L2 speech perception training, the type of CF that has been adopted most frequently has been uniform and simple, conveying a right-or-wrong message (Derwing & Munro, 2015). For instance, when learners choose the right answer, they are provided with a visual and/or audio signal containing the message *right.* When learners choose the wrong answer, they are provided with a visual and/or audio signal indicating *wrong* and also given an opportunity to try again. However, considering that different types of CF are thought to engage different kinds of cognitive and linguistic processing, and that one goal of CF-related studies is to tease apart learning outcomes associated with different types of CF (e.g., Ellis, Loewen, & Erlam, 2006), it is important to investigate whether different types of CF should be adopted and what differential effects they could have during L2 speech perception training. To address these issues, the current study investigated the following research questions:

1. To what extent do Korean learners of L2 English improve their perceptual accuracy in categorizing the English phonemic contrasts /i/-/ɪ/ and /ɛ/-/æ/ in familiar (trained) and unfamiliar (untrained) materials as a result of computer-assisted perception training?
2. To what extent do the training effects differ according to the type of auditory or visual CF (i.e., rejection plus target form, rejection plus nontarget form, rejection plus target and nontarget forms, or *wrong* shown on the computer screen)?

Besides investigating effects of four different types of explicit CF on L2 speech perception, this study also sought to contribute to the debate concerning the importance of providing versus eliciting target L2 form through CF. This study also allowed for examining whether and to what extent learners benefit from CF that provides the target form in the input, from CF that requires learners to retrieve the target form, and from a combination of both CF types. This research thus adds a new dimension to this issue by targeting L2 perception rather than L2 production.

Method

Participants

Participants in this study included 100 Korean learners of L2 English (73 females, 27 males), with a mean age of 30.3 years ($SD = 9.69$). Most participants were students learning English in private or university language institutes located in the Montreal area. The majority had resided in Korea until they were 18 years old; their average length of residence in English-speaking countries, including Canada, was 19.2 months ($SD = 19.1$). The L2 participants self-reported their L2 proficiency as intermediate or advanced based on the length of their learning experience in formal instruction settings ($M =$ 9.8 years, $SD = 8.62$). Participants also included 26 native speakers of English (16 females, 10 males) who recorded the target stimuli ($n = 6$) and served as baseline listeners ($n = 20$). These participants were speakers of North American English from English-speaking provinces in Canada or from the United States, with a mean age of 25.0 ($SD = 5.98$); they had moved to Montreal to pursue undergraduate or graduate studies in English-medium universities.

Procedure

The procedure comprised four steps: (a) a pretest, which also served as a baseline test for the native speakers of English; (b) computer-assisted perception training; (c) an immediate posttest; and (d) a delayed posttest. Before the training sessions, 100 L2 learners and 20 L1 listeners completed a pretest at a research laboratory. The L2 learners were then randomly assigned to one of four treatment groups or the control group ($n = 20$ per group) before participating in eight computer-assisted perception training sessions distributed over a 2-week period. Participants in the four treatment groups received a specific type of CF when they made perceptual errors during the training sessions:

1. a rejection followed by the target form (Target group – auditory);
2. a rejection followed by the nontarget form (Nontarget group – auditory);
3. a rejection followed by target and nontarget forms (Combination group – auditory);
4. *wrong* shown on the computer screen (Wrong group – visual).

After the last training session, the learners completed an immediate posttest and, 2 weeks later, a delayed posttest.

Target Stimuli

The target stimuli were two English vowel contrasts. A number of L2 speech perception training studies have focused on nonnative consonants, yet L2 learners have more difficulty acquiring L2 vowels than consonants, often regardless of their L1 background (Munro & Derwing, 2008; Neri, Cucchiarini, & Strik, 2006). Moreover, considering that vowels are more crucial than consonants in terms of word intelligibility (Bent, Bradlow, & Smith, 2007), L2 vowel training should be a pedagogical priority. Two English vowel contrasts (/i/-/ɪ/ and /ɛ/-/æ/) were selected as targets. Due to the absence of /ɪ/ and /æ/ in their L1, Korean learners of L2 English have difficulty categorizing and discriminating these nonnative contrasts (Baker, Trofimovich, Mack, & Flege, 2002; Flege, Bohn, & Jang, 1997; Ingram & Park, 1997; Tsukada et al., 2005), more so with /ɛ/-/æ/ than with /i/-/ɪ/.

Eighteen sets of English minimal pairs targeted /i/-/ɪ/ and another 18 sets of English minimal pairs targeted /ɛ/-/æ/ (all target materials are listed in Appendix S1 in the Supporting Information online). All word pairs, which followed a monosyllabic consonant-vowel-consonant pattern with various onsets and codas, were selected from the Corpus of Contemporary American English (Davies, 2008). Except for the words with asterisks in Appendix S1, all target words were of high frequency in the corpus (higher than 15 occurrences per million words). The first 12 sets of word pairs for /i/-/ɪ/ and /ɛ/-/æ/ were designated as trained words, whereas the remaining 6 sets were used as untrained words, to determine if learners could transfer phonetic-level knowledge from trained to novel (untrained) words. The trained words appeared in the training and testing sessions, whereas the untrained words were provided during the testing sessions only.

Each word was recorded twice in a carrier phrase "I said . . . " by L1 speakers (three males labelled as M1, M2, and M3, and three females labelled as F1, F2, and F3). The carrier phrase encouraged the speakers to focus narrowly on the target words, all of which were spoken with similar prosody. Speakers M1, M2, F1, and F2 were asked to record both the trained and untrained words. In contrast, Speakers M3 and F3 recorded only the trained words, which were used for testing only, to measure the extent to which learners could categorize the target contrasts in words occurring in the training sessions but spoken in different (i.e., unfamiliar) voices. Using Praat (Boersma & Weenink, 2013), the target words were extracted from the carrier sentences and digitized at 44,100 Hz, with a 16-bit resolution, after which one out of two productions from each speaker was selected. Finally, acoustic analyses targeting vowels (e.g., first and second formant frequencies, duration, and pitch) confirmed that all vowel tokens fell within the ranges of relevant vowel categories based on previous acoustic studies (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995; Yang, 1996).

Training and Testing Sessions

Forced-choice identification training is optimal for training L2 learners (Wang & Munro, 2004), so computer-assisted perception training in this study was implemented as forced-choice identification. The training was programmed with Web-based computer scripts that included JavaScript, PHP, and MySQL (Nixon, 2012). In each training trial, one member of a target word pair was played to participants, with two response options available on the screen. For instance, participants first heard the word *ship*, then saw two orthographic response options (*sheep* and *ship*) and were asked to answer the question "What did s/he say?" by selecting the appropriate response option. The motivation to use orthographic labels complied with previous research, which suggested that L2 learners might benefit from orthographic information during lexical and phonological learning (e.g., Escudero, 2015). There was no predetermined time interval between trials. Because there was a "repeat" button available, L2 learners could listen to each stimulus multiple times until they were sure of their answers. Once they selected their answer, a CF intervention followed, after which the next trial began. In contrast, the next trial was automatically played without any CF in the control group.

If a learner selected the wrong response alternative, different types of CF were provided according to group. Speakers M1 and F1 and Speakers M2 and F2 were paired for the CF intervention. That is, when Speaker M1 provided the target word, Speaker F1 provided the corresponding CF and vice versa. The same method was used for Speakers M2 and F2. For example, if the target word *hit* was heard, but the learner selected *heat*, the following types of CF were provided: "No, s/he said *hit*" (Target group); "No, not *heat*" (Nontarget group); "No, s/he said *hit*, not *heat*" (Combination group); and a word card with *wrong* written in red (Wrong group). In contrast, if a learner chose the correct response alternative, those in the auditory CF groups (i.e., Target, Nontarget, and Combination groups) were all given positive oral confirmation in the form of *yes*, whereas the word *right* appeared in blue on the screen for those in the Wrong group. To help learners notice CF explicitly, a pop-up message saying *okay?* immediately followed the CF interventions; learners were required to click *yes* to move on to the next trial. Although the control group engaged in the same perception training, there was no CF regardless of the response alternative selected. As soon as a response was made, the next trial began.

One training session comprised a total of 384 trials, that is, 48 trained words recorded by Speakers M1, M2, F1, and F2, repeated twice. Including four speakers in the training set was motivated by the fact that three to six talkers are known to yield promising results given the importance of speaker variability during perception training (Thomson, 2011). All trials were randomized for each learner in each training session. It took approximately 1 to 1.5 hours to complete one session, and each learner participated in eight training sessions within 2 weeks.

Forced-choice identification tests designed using Praat (Boersma & Weenink, 2013) were used to measure learners' perceptual accuracy, given the importance of using assessment tasks that are compatible with training tasks (Hardison, 2012). The testing environment was thus similar to the training, that is, a sound file representing one target word was played (e.g., *heat* or *bad*), after which both response alternatives were orthographically shown on the screen (e.g., *hit* and *heat*; *bed* and *bad*), and learners were prompted to select what they heard. In contrast to the perception training, there was no repeat button available during the assessment tasks. All learners thus listened to each stimulus only once. Again, there was no predetermined time interval between trials, and by pressing the *next* button on the screen, learners proceeded from one trial to the next in a self-paced manner.

In the pretest, immediate posttest, and delayed posttest, the learners completed a total of 384 trials, which included trained stimuli spoken in familiar voices (48 words × 4 speakers), trained stimuli spoken in unfamiliar voices (48 words × 2 speakers), and untrained stimuli spoken in familiar voices (24 words × 4 speakers). The baseline test for the 20 native speakers of English consisted of the same 384 trials. The baseline test was set to ensure that the participants relied on their linguistic skills and not on problem-solving skills related to task difficulty. It took approximately 1 hour to complete each test. As in the perception training, all of the trials were randomized for each participant in each testing session. The L2 participants took the pretest 1 day before the first training session, the immediate posttest on the same day of the last training session, and the delayed posttest 2 weeks after the immediate posttest.

Data Analysis

The target scores were based on percentages of correct responses to the target contrasts, which were analyzed statistically with alpha set at .05. A multivariate analysis of variance (MANOVA) and a discriminant function analysis (DFA) were employed to compare pretest, immediate posttest, and delayed posttest scores separately for (a) words occurring in the training sessions (trained words), (b) words occurring in the training sessions recorded in unfamiliar voices (trained words spoken in unfamiliar voices), and (c) words not occurring in the training sessions (untrained words). For between-group contrasts (Plonsky & Oswald, 2014), each of the four CF treatment groups was compared with the control group, and Cohen's d effect size values (Cohen, 1988) were calculated and classified as small ($.40 \leq d < .70$), medium ($.70 \leq d < 1.00$), or large ($d \geq 1.00$). Based on the standardized z score of 3.29, it was confirmed that there were no outliers in the data set. Moreover, statistical assumptions for MANOVA were verified by using a Shapiro-Wilk test for normality of distribution, a positive Pearson correlation between the dependent variables, Levene's test for homogeneity of between-group variances, and Mauchly's test for sphericity of within-group variances. The assumptions of Box's M tests were considered met given that none of the p values were below .001 and that all group sizes were equal (see Field, 2013). Following Field, Pillai's Trace was considered the most powerful statistic to analyze the data, using DFA as a follow-up to a significant MANOVA.[1]

Results

Overall, the participants in the four CF groups received an average of 100.5 CF instances ($SD = 42.6$) per training session (i.e., 384 trials). Specifically, there were on average 99.2 ($SD = 35.2$), 105.2 ($SD = 52.7$), 91.6 ($SD = 45.2$), and 106.1 ($SD = 36.7$) CF instances per session in the Target, Nontarget, Combination, and Wrong conditions, respectively. A one-way ANOVA confirmed that the participants received a similar amount of CF regardless of group, $F(3, 76) = .48, p = .696$.

Trained Words

The participants' performance for words occurring in the training sessions was assessed through a repeated-measures MANOVA conducted with two dependent variables (percent correct responses for /i/-/ɪ/ and /ɛ/-/æ/), group as a between-group independent variable (Target, Nontarget, Combination, Wrong, control), and time as a within-group independent variable (pretest, immediate posttest, delayed posttest). The MANOVA yielded a statistically significant effect for group, $V = .18, F(8, 190) = 2.29, p = .023$, and a statistically significant effect for time, $V = .63$, $F(4, 92) = 39.40$, $p < .001$. In addition, there was a statistically significant group × time interaction, $V = .37$, $F(16, 380) = 2.39$, $p = .002$; this interaction effect was explored further through DFAs. Table 1 summarizes the participants' performance across groups and time.

A DFA, conducted to investigate whether the dependent variables in the MANOVA discriminated among the five groups at the three testing sessions, revealed two functions at the pretest, the first of which accounted for 93.6% of the variance (canonical $R^2 = .038$) and the second accounted for 6.4% of the variance ($R^2 = .003$). A combination of these functions did not significantly

Table 1 Mean percentage accuracy scores and standard deviations (in parentheses) over time by group and target contrast for trained words ($n = 20$ per group)

	Pretest		Immediate posttest		Delayed posttest	
Group	/i/-/ɪ/	/ɛ/-/æ/	/i/-/ɪ/	/ɛ/-/æ/	/i/-/ɪ/	/ɛ/-/æ/
Target	72.3 (11.4)	59.8 (15.9)	81.9 (10.5)	79.0 (11.4)	82.7 (9.4)	76.4 (11.6)
Nontarget	70.5 (15.2)	59.1 (17.4)	79.3 (14.1)	77.4 (10.9)	79.5 (13.4)	74.7 (12.3)
Combination	68.1 (16.4)	57.5 (19.9)	84.2 (10.9)	81.0 (10.7)	80.5 (12.5)	79.1 (11.8)
Wrong	68.1 (8.2)	58.5 (13.2)	80.2 (9.7)	75.5 (13.4)	77.0 (11.3)	75.9 (13.7)
Control	66.2 (16.7)	59.7 (15.9)	67.8 (13.8)	60.7 (10.2)	64.7 (12.5)	59.7 (12.8)

discriminate the groups at the pretest, $\Lambda = .96$, $\chi^2(8) = 3.95$, $p = .862$. In addition, removing the first function revealed that the second function on its own did not significantly discriminate the groups at the pretest, $\Lambda = .99$, $\chi^2(3) = .26$, $p = .968$. All effect sizes for differences between the control group and each of the four CF groups were also small: both for /i/-/ɪ/ accuracy (Target $d = .44$; Nontarget $d = .28$; Combination $d = .12$; Wrong $d = .15$) and for /ɛ/-/æ/ accuracy (Target $d = .01$; Nontarget $d = .04$; Combination $d = .13$; Wrong $d = .08$).

DFAs revealed two functions at the immediate posttest. The first explained 98.7% of the variance ($R^2 = .316$) and the second accounted for 1.3% of the variance ($R^2 = .006$). Unlike the pretest, however, a combination of the functions significantly discriminated the groups at the immediate posttest, $\Lambda = .68$, $\chi^2(8) = 36.81$, $p < .001$. The second function, when evaluated alone, did not reach significance, $\Lambda = .99$, $\chi^2(3) = .58$, $p = .901$. The correlations between the outcomes and the discriminant functions revealed that the /i/-/ɪ/ accuracy rates at the immediate posttest loaded fairly evenly onto both functions ($r = .72$ for the first function and $r = .70$ for the second). However, /ɛ/-/æ/ accuracy rates loaded more strongly on the first function ($r = .96$) than the second function ($r = -.28$). Because the first function accounted for 98.7% of the variance and that the second function was nonsignificant, we focused on the first function with respect to the values of group centroids, which showed that the first function differentiated the four CF treatment groups from the control group (–1.28). In particular, group centroid values were highest in the Combination group and smallest in the Wrong group: Combination (.59), Target (.39), Nontarget (.21), and Wrong (.09). All effect sizes for between-group contrasts (except for the Nontarget-control comparison for /i/-/ɪ/) were large for both /i/-/ɪ/ accuracy (Target $d = 1.18$; Nontarget $d = .85$; Combination

$d = 1.35$; Wrong $d = 1.07$) and /ɛ/-/æ/ accuracy (Target $d = 1.74$; Nontarget $d = 1.62$; Combination $d = 1.99$; Wrong $d = 1.28$).

DFAs identified two functions at the delayed posttest, the first of which accounted for 92.5% of the variance ($R^2 = .283$) and the second explained 7.5% of the variance ($R^2 = .031$). As in the immediate posttest, a combination of the two functions significantly discriminated the groups, $\Lambda = .70$, $\chi^2(8) = 34.70$, $p < .001$, whereas the second function alone did not discriminate the groups, $\Lambda = .96$, $\chi^2(3) = 2.99$, $p = .393$. The correlations between the outcomes and the discriminant functions revealed that the /i/-/ɪ/ accuracy rates at the delayed posttest loaded onto both functions ($r = .86$ for the first function and $r = .51$ for the second) but that the /ɛ/-/æ/ accuracy rates loaded more strongly on the first function ($r = .89$) than the second function ($r = -.45$). As in the immediate posttest, the values of group centroids indicated that the first function differentiated the four CF groups from the control group (–1.20). Again, group centroid values were highest in the Combination group and lowest in the Wrong group: Combination (.45), Target (.42), Nontarget (.19), and Wrong (.13). All effect sizes for contrasts between each of the four CF groups and the control group were large, for both /i/-/ɪ/ accuracy (Target $d = 1.67$; Nontarget $d = 1.17$; Combination $d = 1.30$; Wrong $d = 1.06$) and /ɛ/-/æ/ accuracy (Target $d = 1.40$; Nontarget $d = 1.23$; Combination $d = 1.62$; Wrong $d = 1.25$).

In summary, all groups showed similar perceptual accuracy at the pretest. However, the four CF groups significantly outperformed the control group at the immediate and delayed posttests, with perceptual accuracy increasing in the following order: Wrong group, Nontarget group, Target group, and Combination group. As expected, the native speakers of English showed ceiling effects for both /i/-/ɪ/ accuracy ($M = 99.9$, $SD = .47$) and /ɛ/-/æ/ accuracy ($M = 96.9$, $SD = 2.48$).

Trained Words Spoken in Unfamiliar Voices

Similar analyses were carried out to examine the perceptual accuracy across the five groups for trained words spoken in unfamiliar voices. A repeated-measures MANOVA revealed a statistically significant effect for group, $V = .16$, $F(8, 190) = 2.07$, $p = .041$, a statistically significant effect for time, $V = .76$, $F(4, 92) = 73.68$, $p < .001$, and a statistically significant group × time interaction, $V = .38$, $F(16, 380) = 2.47$, $p = .001$, which was explored further by means of DFAs. Table 2 summarizes the participants' performance across groups and time.

At the pretest, a DFA found two functions, one explaining 84.1% of the variance ($R^2 = .057$) and the other accounting for 15.9% of the variance

Table 2 Mean percentage accuracy scores and standard deviations (in parentheses) over time by group and target contrast for trained words spoken in unfamiliar voices ($n = 20$ per group)

	Pretest		Immediate posttest		Delayed posttest	
Group	/i/-/ɪ/	/ɛ/-/æ/	/i/-/ɪ/	/ɛ/-/æ/	/i/-/ɪ/	/ɛ/-/æ/
Target	72.4 (8.7)	62.0 (8.9)	82.5 (9.0)	79.3 (8.2)	83.9 (7.5)	82.5 (11.1)
Nontarget	68.1 (14.9)	60.5 (14.9)	79.6 (13.7)	76.1 (10.7)	79.9 (14.7)	77.9 (10.1)
Combination	67.7 (9.1)	53.9 (17.4)	82.7 (7.4)	81.9 (9.8)	81.8 (7.6)	80.2 (9.5)
Wrong	71.8 (10.7)	61.0 (15.5)	80.8 (11.0)	73.4 (18.2)	77.9 (11.0)	77.4 (13.0)
Control	67.9 (12.2)	56.9 (14.9)	71.6 (16.7)	63.9 (18.9)	70.5 (14.9)	67.1 (10.7)

(R^2 = .011). A combination of the two functions did not significantly discriminate the groups, Λ = .93, $\chi^2(8) = 6.67$, p = .573, nor did the second function on its own, Λ = .99, $\chi^2(3) = 1.08$, p = .782. All effect sizes for the contrasts between the control group and each CF group were also small, for both /i/-/ɪ/ accuracy (Target d = .44; Nontarget d = .02; Combination d = .02; Wrong d = .35) and /ɛ/-/æ/ accuracy (Target d = .43; Nontarget d = .25; Combination d = .19; Wrong d = .28).

At the immediate posttest, the analysis revealed two functions: The first explained 96.3% of the variance (R^2 = .207) whereas the second accounted for 3.7% of the variance (R^2 = .010). A combination of the two functions significantly discriminated the groups, Λ = .79, $\chi^2(8) = 23.08$, p = .003, but the second function on its own did not, Λ = .99, $\chi^2(3) = .96$, p = .811. In particular, the structure matrix showed that the /i/-/ɪ/ accuracy rates at the immediate posttest loaded similarly onto the first function (r = .67) and the second function (r = .74). However, the /ɛ/-/æ/ accuracy rates loaded more strongly on the first function (r = .89) than the second function (r = –.45). The first function significantly differentiated the auditory CF groups from the Wrong group (–.03) and the control group (–.92), with group centroid values being highest in the Combination group (.51), intermediate in the Target group (.36), and smallest in the Nontarget group (.07). Similarly, the effect sizes were small for contrasts between the Wrong group and the control group, for both /i/-/ɪ/ accuracy (Wrong d = .67) and /ɛ/-/æ/ accuracy (Wrong d = .53). However, effect sizes were small to large for contrasts between each auditory CF group and the control group: for /i/-/ɪ/ accuracy (Target d = .83; Nontarget d = .54; Combination d = .88) and /ɛ/-/æ/ accuracy (Target d = 1.09; Nontarget d = .82; Combination d = 1.23).

At the delayed posttest, a DFA yielded two functions, the first accounting for 99% of the variance (R^2 = .227) and the second accounting for 1% (R^2 = .003). As in the immediate posttest, a combination of the functions significantly discriminated the groups, $\Lambda = .77$, $\chi^2(8) = 24.79$, $p = .002$, whereas the second function by itself did not, $\Lambda = .99$, $\chi^2(3) = .28$, $p = .964$. The /i/-/ɪ/ accuracy rates at the delayed posttest loaded similarly onto the first function ($r = .74$) and the second function ($r = .67$), but the /ɛ/-/æ/ accuracy rates loaded more strongly on the first function ($r = .92$) than the second function ($r = –.40$). As in the immediate posttest, the first function significantly distinguished the auditory CF groups from the Wrong group (–.01) and the control group (–.98), with group centroid values being highest in the Target group (.56), intermediate in the Combination group (.33), and lowest in the Nontarget group (.10). The effect sizes for contrasts between the Wrong group and the control group were small to medium for both /i/-/ɪ/ accuracy (Wrong $d = .58$) and /ɛ/-/æ/ accuracy (Wrong $d = .89$); effect sizes were mostly medium to large for contrasts between each auditory CF group and the control group: for /i/-/ɪ/ accuracy (Target $d = 1.16$; Nontarget $d = .65$; Combination $d = .98$) and /ɛ/-/æ/ accuracy (Target $d = 1.45$; Nontarget $d = 1.07$; Combination $d = 1.33$).

In sum, at the pretest, all groups showed similar perceptual accuracy for trained words spoken in unfamiliar voices. However, at the immediate and delayed posttests, the auditory CF groups outperformed the Wrong group and the control group. Accuracy was highest in the Combination group and lowest in the Nontarget group at the immediate posttest; accuracy was highest in the Target group and lowest in the Nontarget group at the delayed posttest. The native speakers of English revealed ceiling effects for both /i/-/ɪ/ accuracy ($M = 99.6$, $SD = .85$) and /ɛ/-/æ/ accuracy ($M = 98.4$, $SD = 1.89$).

Untrained Words

To compare the groups' performance accuracy for untrained words, similar analyses were carried out. A repeated-measures MANOVA revealed a statistically significant effect for time, $V = .59$, $F(4, 92) = 33.54$, $p < .001$, but no significant effect for group, $V = .15$, $F(8, 190) = 1.87$, $p = .067$, or a significant group × time interaction, $V = .24$, $F(16, 380) = 1.48$, $p = .103$. Because a significant time effect (independent of group effects) was not of primary interest, no follow-up analyses were conducted; instead, we focused on reporting between-group effect sizes. Table 3 summarizes the participants' performance across groups and time.

Although the group × time interaction failed to reach significance, effect size values showed trends similar to those in previous analyses. For instance,

Table 3 Mean percentage accuracy scores and standard deviations (in parentheses) over time by group and target contrast for untrained words ($n = 20$ per group)

	Pretest		Immediate posttest		Delayed posttest	
Group	/i/-/ɪ/	/ɛ/-/æ/	/i/-/ɪ/	/ɛ/-/æ/	/i/-/ɪ/	/ɛ/-/æ/
Target	69.0 (13.3)	56.9 (10.0)	80.6 (12.0)	69.8 (13.3)	80.6 (10.9)	68.5 (10.3)
Nontarget	64.6 (13.7)	56.5 (12.8)	75.0 (13.1)	66.7 (12.7)	77.1 (11.9)	65.0 (13.9)
Combination	66.3 (15.8)	58.1 (15.0)	78.1 (12.5)	67.3 (12.0)	79.8 (12.1)	66.5 (12.3)
Wrong	68.1 (8.3)	57.5 (9.6)	80.4 (11.4)	68.8 (13.9)	79.2 (10.7)	64.4 (11.7)
Control	63.5 (11.5)	53.9 (10.4)	62.8 (17.5)	56.8 (10.1)	68.5 (18.4)	59.5 (13.5)

all effect sizes for contrasts between the control group and each CF group were small at the pretest: for /i/-/ɪ/ accuracy (Target d = .45; Nontarget d = .09; Combination d = .21; Wrong d = .47) and /ɛ/-/æ/ accuracy (Target d = .30; Nontarget d = .23; Combination d = .34; Wrong d = .37). However, at the immediate posttest, medium-to-large effect sizes were found for both /i/-/ɪ/ accuracy (Target d = 1.22; Nontarget d = .81; Combination d = 1.03; Wrong d = 1.22) and /ɛ/-/æ/ accuracy (Target d = 1.13; Nontarget d = .88; Combination d = .97; Wrong d = 1.01). At the delayed posttest, small-to-medium effects were observed for contrasts between the control group and each CF group: for /i/-/ɪ/ accuracy (Target d = .82; Nontarget d = .57; Combination d = .74; Wrong d = .73) and /ɛ/-/æ/ accuracy (Target d = .77; Nontarget d = .41; Combination d = .55; Wrong d = .40). Again, the native speakers of English showed near-perfect perceptual accuracy for /i/-/ɪ/ (M = 99.8, SD = .93) and /ɛ/-/æ/ (M = 96.7, SD = 4.40).

Discussion

The four CF groups outperformed the control group at the immediate and delayed posttests in their perception of trained words. The auditory CF groups also outperformed the Wrong group and the control group at the two posttests in their perception of trained words spoken in unfamiliar voices. There were no significant group differences between each CF group and the control group at the immediate and delayed posttests in their perception of untrained words; nonetheless, descriptive statistics and effect size analyses showed that the CF groups increased their scores across all tests relative to the control group. Among the CF groups, the auditory CF groups fared better than the visual (Wrong) group, which did not outperform the control group for trained words spoken in unfamiliar voices. And within the auditory CF groups, all significant

DFA analyses indicated that the Combination group showed the highest performance at the immediate and delayed posttests, except for trained words spoken in unfamiliar voices at the delayed posttest. In light of these results, the benefits of auditory CF types—and especially CF which combines both target and nontarget forms—will be discussed.

Benefits of Auditory CF for L2 Speech Perception Training

In terms of informing L2 learners that their linguistic knowledge was not accurate, the auditory CF and visual conditions were all relatively explicit: The feedback in all auditory CF conditions began with *no*, and visual CF featured the word w*rong*. What differentiated auditory and visual CF was the amount of linguistic information provided. The visual condition was successful with respect to rejecting learners' responses, but provided no other acquisitional aids that might help them to restructure their ill-formed knowledge. Accordingly, the learners in the Wrong group fell short of outperforming even the control group for the trained words spoken in unfamiliar voices. This finding suggests that visual CF feedback which includes simply an acknowledgement of the error (*wrong*) has a limited impact on acquiring the target phonemes.

In contrast, the auditory types of CF gave learners other acquisitional aids following the verbal rejection *no*. For example, the Target type of CF entailed positive input: By hearing a positive exemplar following their perceptual error, learners were likely induced to notice the gap between the target form and their perceptual error (see Schmidt & Frota, 1986). As for the Nontarget type of CF, it likely pushed learners to find the right form relative to their perceptual error. Regarding the Combination condition, a combination of the two CF types was found to result in a joint positive effect, which we will discuss next.

Benefits of Combination Type of CF

According to PAM, Korean learners of L2 English are likely to assimilate the English phonemes /ɪ/ and /æ/ to the existing Korean phonemes /i/ and /ɛ/, respectively; this would make learners desensitized to the phonetic differences between /i/ and /ɪ/ and between /ɛ/ and /æ/ (see Baker et al., 2002). The Target and Nontarget CF conditions each provided learners with a different type of information concerning the contrasts, that is, either a positive exemplar in the Target condition or a verbatim repetition of the erroneous response in the Nontarget condition. By providing the learners with a positive exemplar, the Target condition reinforced the target pattern. In the Nontarget condition, the learners were drawn to notice the nontarget pattern, which likely encouraged them to consider the alternative. Considering that Korean learners of L2 English have difficulty categorizing the target vowels because they are desensitized to the phonetic differences between /i/ and /ɪ/ and between /ɛ/ and /æ/, the effects of providing only one or the other type of information (i.e., reinforcing either the target or nontarget form) were not as robust as those of providing both types of information in the Combination CF condition. By juxtaposing the target form with the nontarget form, and thus optimizing awareness of the phonetic differences (i.e., psychoacoustic salience), the Combination condition proved most effective. The learners in this condition were more readily able to compare the target word and their erroneous response without the burden of having to retrieve either their erroneous response or the target word from memory, which was likely the case in the Target and Nontarget conditions. The additional support provided to learners in the Combination condition thus likely reduced their cognitive processing load in a way that resulted in higher performance in categorizing the target vowels in comparison to the other auditory CF conditions.

Finally, Hardison (2012), among others, argued that successful L2 speech perception training should show generalization to new stimuli, such as novel voices or words. We found some positive CF effects for unfamiliar (novel) voices but no significant CF effects for untrained words. Descriptive statistics and effect size analyses, however, showed that the CF groups improved their performance across all tests relative to the control group. These findings raise the question as to whether the nature of the knowledge gained as a result of the perception training was word specific. We thus speculate that eight training sessions might not be sufficient for L2 learners to acquire phonetic-level knowledge that extends beyond the words experienced during training.

Conclusion and Future Directions

Although several CF studies have investigated whether L2 learners benefit more from hearing the target form in the input or from having to retrieve it from their own resources, Lyster et al. (2013) concluded their review of CF research with a call for research into combinations of different CF types. In response to this call, the current study revealed that providing both target and nontarget forms improved L2 learners' perceptual accuracy with respect to L2-specific phonemic contrasts, likely by increasing learners' extent of noticing relevant phonetic differences between both members of each target contrast. We suggest that the Combination condition was effective because it reduced learners' cognitive burden of having to retrieve an erroneous response from memory in the Target condition and to retrieve the target word from memory in the Nontarget condition. Additionally, the *wrong* visual type of CF, which has been the most frequent in speech perception training, was the least effective at helping L2 learners accurately identify members of difficult L2 phonemic contrasts and generalize results of training to the trained words spoken in unfamiliar voices.

As for future directions, it would be of interest to investigate the extent to which L2 learners need to be exposed to perception training in order for its effects to extend to untrained words. In addition, considering that, in this study, learners could listen to each stimulus repeatedly throughout the training, it can be suggested that high repetition of input during training might influence learners' performance regardless of CF type. Therefore, to gain a clearer understanding of differential effects of various CF types, it would be important to replicate this study, controlling the amount of repeated experience with training materials during L2 speech perception training. Moreover, it would be interesting to explore the extent to which individual variables (e.g., age, language aptitude, proficiency, and length of residence and instruction) influence perceptual accuracy, beyond the effects of L2 speech training. Finally, most L2 learners encounter their target language(s) through classroom-based instruction rather than in computer-assisted training. In this sense, this study raises other questions concerning the feasibility and effectiveness of implementing similar perception-based training in classroom settings. Arguably, in addition to focusing on laboratory-based perception training in the form of decontextualized practice, it would be highly important to target classroom-based training conducted in the context of meaningful activities. We hope that these suggestions will serve as catalysts for future research in the fields of CF and L2 speech perception.

Final revised version accepted 31 August 2015

Note

1 Although the default approach to explore significant MANOVA effects is to carry out individual univariate ANOVAs for each dependent variable, according to Field (2013), this approach weakens the linear combination of the MANOVA dependent variables while also increasing the likelihood of Type I error. Therefore, DFA, which is mathematically the inverse of a MANOVA, is preferred. DFA suggests how dependent variables in MANOVA discriminate a grouping variable; it works by identifying linear variates that best differentiate the groups and these linear variates are the functions in the DFA.

References

Baese-Berk, M. M. (2010). *An examination of the relationship between speech perception and production*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002). The effect of perceived phonetic similarity on non-native sound learning by children and adults. In S. A. Fisch, B. Scarabela, & A.-H. Do (Eds.), *Proceedings of the 26th Boston University Conference on Language Development* (Vol. *26*, pp. 36–47). Somerville, MA: Cascadilla Press.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Phonemic errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 331–347). Amsterdam: John Benjamins. doi:10.1075/lllt.17.28ben

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins. doi:10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program] (Version 5.3.41). Retrieved from http://www.praat.org

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning*, *33*, 499–526. doi:10.1111/j.1467-1770.1983.tb00946.x

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310. doi:10.1121/1.418276

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251. doi:10.1017/S0952675700001019

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

Davies, M. (2008). *The corpus of contemporary American English (COCA): 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca

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

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*, 476–490. doi:10.1017/S026144480800551X

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins. doi:10.1075/lllt.42

Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, *41*, 25–37. doi:10.1016/j.system.2013.01.005

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–368. doi:10.1017/S0272263106060141

Escudero, P. (2006). Second language phonology: The role of perception. In M. Pennington (Ed.), *Phonology in context* (pp. 109–134). New York: Palgrave Macmillan. doi:10.1057/9780230625396

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, *36*, 7–22. doi:10.1017/S0142716414 00040X

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London: Sage.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399–423. doi:10.2307/3588487

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (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. 319–355). Berlin, Germany: 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. doi:10.1006/jpho.1997.0052

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*, 495–522. doi:10.1017/S0142716403000250

Hardison, D. M. (2012). Second language speech perception: A cross-disciplinary perspective on challenges and accomplishments. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 349–363). New York: Routledge.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111. doi:10.1121/1.411872

Ingram, J. C., & Park, S. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343–370. doi:10.1006/jpho.1997.0048

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*, 3267–3278. doi:10.1121/1.2062307

Jamieson, D. G., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, *16*, 201–210.

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research*, *19*, 254–275. doi:10.1177/1362168814541735

Lee, A. H., & Lyster, R. (2015). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*, 1–30. doi:10.1017/S0272263115000194

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 369–377. doi:10.2307/3588485

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365. doi:10.1111/j.1467-9922.2010.00561.x

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255. doi:10.1121/1.408177

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 92–99). Stockholm, Sweden: Royal Institute of Technology/Stockholm University.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886. doi:10.1121/1.1894649

Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, *59*, 453–498. doi:10.1111/j.1467-9922.2009.00512.x

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA. *Studies in Second Language Acquisition*, *32*, 265–302. doi:10.1017/S0272263109990520

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*, 1–40. doi:10.1017/S0261444812000365

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. Catherine-Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Boston University Conference on Language Development* (Vol. *2*, pp. 522–533). Somerville, MA: Cascadilla Press.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us. In A.-H. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Boston University Conference on Language Development* (Vol. *2*, pp. 480–490). Somerville, MA: Cascadilla Press.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior*, *77*, 657–662. doi:10.1016/S0031-9384(02)00916-2

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*, 479–502. doi:10.1111/j.1467-9922.2008.00448.x

Neri, A., Cucchiarini, C., & Strik, H. (2006). Selecting segmental errors in non-native Dutch for optimal pronunciation training. *International Review of Applied Linguistics in Language Teaching*, *44*, 357–404. doi:10.1515/IRAL.2006.016

Nixon, R. (2012). *Learning PHP, MySQL, JavaScript, and CSS: A step-by-step guide to creating dynamic websites* (2nd ed.). Sebastopol, CA: O'Reilly Media.

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

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning*, *44*, 681–734. doi:10.1111/j.1467-1770.1994.tb00635.x

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 379–410). Timonium, MD: York Press.

Rvachew, S., & Jamieson, D. G. (1995). Learning new speech contrasts: Evidence from adults learning a second language and children with speech disorders. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 411–432). Timonium, MD: York Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928. doi:10.1126/science.274.5294.1926

Saito, K. (2013a). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning*, *63*, 499–529. doi:10.1111/lang.12015

Saito, K. (2013b). Re-examining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, *35*, 1–29. doi:10.1017/S0272263112000666

Saito, K., & Lyster, R. (2012a). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*, 595–633. doi:10.1111/j.1467-9922.2011.00639.x

Saito, K., & Lyster, R. (2012b). Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly*, *46*, 387–398. doi:10.1002/tesq.25

Saito, K., & Wu, X. (2014). Communicative focus on form and L2 suprasegmental learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition*, *36*, 647–680. doi:10.1017/S027226311400 0114

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Rowley, MA: Newbury House.

Setter, J., & Jenkins, J. (2005). State-of-the-art review article: Pronunciation. *Language Teaching*, *38*, 1–17. doi:10.1017/S026144480500251X

Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. *2*, pp. 593–610). New York: Routledge.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*, 131–145. doi:10.3758/BF03202673

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Amsterdam: John Benjamins. doi:10.1075/sibil.36.09str

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*, 744–765. doi:10.1558/cj.v28i3.744-765

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*, 263–290. doi:10.1016/j.wocn.2004.10.002

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication*, *49*, 2–7. doi:10.1016/j.specom.2006.10.003

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*, 539–552. doi:10.1016/j.system.2004.09.011

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033–1043. doi:10.1121/1.1531176

Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*, *16*, 978–987. doi:10.1162/0898929041502616

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, *24*, 245–261. doi:10.1006/jpho.1996.0013

Supporting Information

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

Appendix S1. Target Stimuli.