CrossMark

Studies in Second Language Acquisition, 2016, **38**, 35–64.
doi:10.1017/S0272263115000194

THE EFFECTS OF CORRECTIVE FEEDBACK ON INSTRUCTED L2 SPEECH PERCEPTION

Andrew H. Lee and Roy Lyster
McGill University

To what extent do second language (L2) learners benefit from instruction that includes corrective feedback (CF) on L2 speech perception? This article addresses this question by reporting the results of a classroom-based experimental study conducted with 32 young adult Korean learners of English. An instruction-only group and an instruction + CF group were exposed to five 1-hr form-focused lessons that drew learners' attention to the nonnative phonemic contrast /i/-/ɪ/, but only the instruction + CF group was given relevant feedback. Forced-choice identification tasks were completed by participants in a pretest, an immediate posttest, and a delayed posttest. The two groups showed similar accuracy on the pretest; however, the instruction + CF group outperformed the instruction-only group on the immediate and delayed posttests as well as on unfamiliar words. The significant predictors for these differences turned out to be perceptual accuracy vis-à-vis /ɪ/-natural and /ɪ/-synthesized sounds. These findings are discussed in terms of the pivotal role played by CF in developing accuracy in L2 speech perception.

Studies of second language (L2) perception training have been conducted exclusively in laboratory settings. However, a close look at how

This research was supported by a Social Sciences and Humanities Research Council of Canada grant (no. 410-2011-0671) awarded to Roy Lyster and the Ellen Edith Grubb Stansfield Fellowship awarded by the McGill Faculty of Education to Andrew H. Lee. Earlier versions of this article were presented at the 2014 annual meeting of the American Association for Applied Linguistics held in Portland, Oregon, and the New Sounds 2013 Conference held in Montreal, Quebec. We gratefully acknowledge all participants and extend special thanks to the three instructors in addition to the anonymous reviewers for providing helpful feedback on earlier drafts.

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

© Cambridge University Press 2015

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

the laboratory perception training has been operationalized in these studies reveals several teaching techniques used in L2 classrooms. One of them is to provide L2 learners with feedback to indicate errors and offer opportunities for correction. Laboratory-based studies have brought to the fore the role played by feedback in the training sessions (Hardison, 2003; Logan, Lively, & Pisoni, 1991; X. Wang & Munro, 2004), which has given rise to the main question driving the current study: What is the role of corrective feedback (CF) in L2 speech perception and training?

Over the past 20 years, many research studies have investigated CF. In terms of phonological targets, however, there are only a few studies, and most of them focus on the development of L2 speech production and the effects of CF on learners' misarticulation (e.g., Saito, 2012; Saito & Lyster, 2012a, 2012b). Yet speech production and perception are inextricably linked. Many researchers (Escudero, 2006; Flege, 1995; Leather, 1999; Llisterri, 1995) have emphasized the role of speech perception in the acquisition of L2 phonology, arguing that speech perception precedes production. In addition, previous findings suggest that L2 learners benefit solely from perception training without explicit pronunciation instruction on L2 production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Jamieson & Rvachew, 1992; Rochet, 1995; Rvachew & Jamieson, 1995; Y. Wang, Jongman, & Sereno, 2003). It is thus speculated that CF may affect L2 learners' perception first and may then trigger successful production of L2 sounds.

With respect to the well-known difficulty in the acquisition of the liquid consonants /ɹ/ and /l/ by Japanese learners of English, even after extensive conversational instruction, Strange and Dittmann (1984) suggested that "improved laboratory training tasks may be useful in establishing categorical perception of these contrasts" (p. 131). Arguably, however, it is more likely for learners to be exposed to their L2 in classroom settings than in laboratory sessions. Given the wide disparity between laboratory-based perception training and instructor-student interaction, it is timely to delve into an exploration of the extent to which learners benefit from instruction and CF on L2 speech perception in simulated classrooms. Accordingly, this study explores the role of L2 speech perception in L2 phonological development and suggests a pivotal role for CF in L2 speech perception training. The current study is thus expected to provide useful support to the fields of L2 education and speech perception research.

BACKGROUND

Linguistic Targets: /i/ versus /ɪ/

It is well known that L2 learners have difficulty perceiving nonnative phonemes. For example, English /i/ (as in *feel* /fil/) and /ɪ/ (as in *fill* /fɪl/)

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

have different spectral information (first formant [F1] is approximately 342–437 Hz for /i/ and approximately 427–483 Hz for /ɪ/) and durational information (approximately 243–306 ms for /i/ and approximately 192–237 ms for /ɪ/; Hillenbrand, Getty, Clark, & Wheeler, 1995). Native speakers of English use both spectral and durational information to categorize these vowels (Bohn & Flege, 1990; Escudero & Boersma, 2004). However, Flege, Bohn, and Jang (1997) found that Korean learners of English could not discern /i/ and /ɪ/ as could native speakers of English. Baker, Trofimovich, Mack, and Flege (2002) found that Korean learners of English tended to categorize /i/ and /ɪ/ as a single phoneme /i/, and Tsukada et al. (2005) also showed that Korean learners of English perceived /ɪ/ as /i/, /e/, or /ɛ/, whereas /i/ was always categorized as /i/.

In sum, many empirical studies have indicated that Korean learners of English have difficulty categorizing /i/ and /ɪ/ and often classify both of these sounds as the phoneme /i/. Such empirical findings are also supported by several speech perception models. According to the perceptual assimilation model (PAM; Best, 1995; PAM-L2, Best & Tyler, 2007), the first language (L1)-specific perception system impedes L2 learners from detecting patterns of articulatory gestures in nonnative sounds because nonnative sounds are assimilated into phonemes pervading the L1-specific speech perception system. In particular, if two nonnative segments are assimilated to one L1 category and both are equally compatible with the L1 category (i.e., single-category type), L2 listeners are highly likely to be poor at discerning those nonnative sounds. If both sounds happen to be assimilated to the same L1 segment but the L1 segment is more perceptually similar to one or the other (i.e., category-goodness type), L2 listeners are less likely to have difficulty discriminating them; in this circumstance, the nonnative segment that is closer to the L1 category is prioritized. Given that /i/ and /ɪ/ exist in the English vowel inventory, whereas only /i/ exists in the Korean vowel inventory (Lee, 1993), Korean learners of English are predicted to show a single-category or category-goodness assimilation pattern in that they are unlikely to demonstrate excellent categorization with respect to these L2 vowels.

The speech learning model (SLM; Flege, 1995) postulates that L2 speech learning is constrained by several factors, one of which is perceived crosslinguistic similarity. When L2 learners acquire nonnative phonemes, the phonemes are classified as either new or similar. Simply put, L2 learners have more difficulty perceiving similar sounds than new sounds, and they need more time and intervention to acquire similar sounds compared to new sounds. Considering English /i/ and /ɪ/ within the framework of the SLM, these sounds are rather similar to /i/ in the Korean vowel inventory, and thus Korean learners of English are expected to have difficulty perceiving these sounds correctly. It is also

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

predicted that a great deal of time and learning will be required for Korean learners of English to acquire this phonemic contrast.

The natural referent vowel framework (NRV; Polka & Bohn, 2011) provides an acoustic perspective to explain why Korean learners of English have difficulty perceiving English /ɪ/ compared to /i/. Polka and Bohn found directional asymmetries occurring in infants' vowel perception: Infants easily detected a vowel change from a more central vowel (e.g., /ɪ/) to a more peripheral vowel (e.g., /i/), whereas they were unlikely to do so in the reverse situation (i.e., from /i/ to /ɪ/). The NRV framework explains that such directional asymmetries result from extreme articulatory-acoustic properties of peripheral vowels (peripheral in the vowel space), which can be defined as natural referent vowels. Polka and Bohn (2011) argued that the natural referent vowels are perceptually salient and stable due to formant frequency convergence or focalization (i.e., the convergence between two consecutive formants in a vowel frequency) of peripheral vowels. If formants move close together in frequency, there is an increase in acoustic energy that reinforces the amplitude of each formant. Focalization thus results in perceptual salience (J. L. Schwartz, Abry, Boë, Ménard, & Vallée, 2005).

The most interesting prediction of the NRV framework is that

> this bias may also re-appear when perceivers are mapping out a new vowel system to learn a second language, possibly when new L2 vowel categories fall in regions of the vowel space that are not firmly committed to the L1. (Polka & Bohn, 2011, p. 474)

Therefore, when Korean learners of English are exposed to /i/ and /ɪ/, it is predicted that they perceive /i/ as more salient than /ɪ/. Consequently, this perceptual bias may prevent Korean learners of English from perceiving /ɪ/ and the /i/-/ɪ/ phonemic contrast and, in turn, may delay acquisition of the nonnative phonemic contrast that is crucial in the target language.

Korean learners of English are also known to rely on durational information. Regardless of spectral information (mainly, the F1), if the duration of the vowel is long, they tend to consider it /i/; if the duration of the vowel is short, they classify it as /ɪ/ (Flege et al., 1997). The overuse of durational cues can be explained by Bohn's (1995) desensitization hypothesis and Escudero and Boersma's (2004) evidence of full access (B. D. Schwartz & Sprouse, 1996). Bohn's (1995) desensitization hypothesis posits that, with respect to categorizing contrastive phonemes, if L2 learners are desensitized to a spectral difference between them, durational information is inadvertently entrenched. This is because "duration cues in vowel perception are easy to access whether or not listeners have had specific linguistic experience with them" (Bohn, 1995, p. 294). He claimed that the desensitization hypothesis is closely

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

associated with a universal availability of durational information as a means of categorizing /i/ and /ɪ/.

In contrast, Escudero and Boersma (2004) attributed the overuse of durational cues to evidence of full access (B. D. Schwartz & Sprouse, 1996). By citing distributional learning from the research study by Maye, Werker, and Gerken (2002), Escudero and Boersma argued that a length contrast is an example of category formation (e.g., /i/ vs. /i:/) that can be easily found as a L1 acquisition strategy. However, splitting one assimilated sound is a case of category split (e.g., /i/ vs. /ɪ/) that can hardly be permitted in typical L1 acquisition. In sum, they concluded that L2 acquisition entails access to a L1-like acquisition device (i.e., full access) wherein L2 learners prefer to use length differences rather than to split one category into two different categories to make a contrast.

The Effects of Perception Training

In light of the perceptual confusion on the part of Korean learners of English, the next question is the following: In spite of the difficulty, is it possible for Korean learners of English to improve their perceptual accuracy? To delve into this question, a number of empirical studies have been conducted on the effects of perception training. These studies have established that learners benefit from training in developing perceptual accuracy, especially concerning consonants; however, a relatively smaller number of findings pertain to vowels (e.g., Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Lee, 2008; Thomson, 2007; X. Wang, 2002; X. Wang & Munro, 2004). Among these, X. Wang and Munro (2004) investigated how Mandarin and Cantonese learners of English perceive /i/ and /ɪ/ and whether they could benefit from computer-based perception training. Their study showed that a group exposed to perception training outperformed a control group in discerning /i/ and /ɪ/ and was also able to transfer this knowledge to new contexts. In a similar study, Lee (2008) found comparable findings with Korean learners of English.

However, a vast number of studies have been conducted in laboratory settings in which L2 learners practice test items via forced-choice identification tasks, A(B)X discrimination tasks, or goodness judgment tasks used for measuring their perceptual accuracy. During training sessions, learners are usually provided with feedback only when they are wrong or when they are both right and wrong. In terms of feedback, when learners choose a right answer, visual and/or audio confirmation is provided. If learners select a wrong answer, visual and/or audio signals are provided to help learners to notice errors and select right answers by repetition (e.g., visual feedback; Leather, 1990).

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Notwithstanding the effects of perception training with feedback, it is unclear whether such laboratory training procedures are compatible with L2 classroom instruction and instructor-student interaction. That is, learners of English as a foreign language (EFL) are most often in classroom settings, which, unlike laboratory settings, entail various combinations of explicit instruction, metalinguistic explanation, exposure to instructional input, engagement with learning tasks, and opportunities to reflect on their learning outcomes. In addition, the types of feedback adopted in laboratory-based training are usually limited, simply conveying a right or wrong message. In contrast, learners in classroom settings are more likely to encounter instances of feedback that are semantically richer than those provided in laboratory settings.

Saito (2013) recently found effects for instruction on the perception of /ɹ/ by Japanese learners of English. Saito and Wu (2014) also investigated the effects of instruction and CF on the perception of Mandarin tones with Cantonese learners of Mandarin. However, in both studies, the form-focused instructional treatment, which included CF, was designed to target L2 learners' production errors, not their perceptual errors. These studies were designed not to directly address the effects of classroom-based perception training but rather to address the transfer effects of production training on perceptual ability. To the best of our knowledge, there is no specific classroom-based research that has investigated whether L2 learners benefit from classroom-based perception training including L2 instruction and oral feedback provided during instructor-student interaction.

Many laboratory-based studies have paid attention to the importance of feedback, albeit simply (e.g., right vs. wrong). For example, Logan et al. (1991) and Hardison (2003) emphasized immediate feedback, mentioning that feedback helped learners to consistently focus on a key feature of the target stimuli. Even in laboratory-based perception training, however, there are few research studies that have investigated the effects of feedback on speech perception. Concluding that "identification training with feedback can improve ESL [English as a second language] speakers' performance on English vowel contrasts," X. Wang and Munro (2004) qualified their findings as "a research focus that so far has received little attention" (p. 550).

Whereas an impressive number of research studies have been published since SLA researchers started to explore the role of CF in L2 learning, there are only a handful of research studies probing the effects of CF on phonological targets, and all of these focus on L2 speech production. Finally, most studies concerning L2 speech perception training were conducted in laboratory settings with computers. Conducted in simulated classroom settings with EFL students and instructors, the current study aims to add to our understanding of the instructional

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

characteristics and effects of perception training in contexts other than individualized tutoring in laboratory settings.

Research Questions

In light of previous research studies and the motivation for the current study, the research questions are as follows:

1. As a result of classroom perception training, to what extent do Korean learners of English improve their perceptual accuracy in categorizing /i/ and /ɪ/ in words occurring in the training sessions as well as in words not occurring in the training sessions?
2. To what extent do the training effects differ between a group receiving instruction only and a group receiving instruction plus CF?

METHODOLOGY

Participants

Participants in this study included a total of 32 Korean learners of EFL (instruction only: $n = 16$; instruction + CF: $n = 16$) and 21 L1 English participants (L1 speakers: $n = 2$; L1 listeners: $n = 16$; ESL instructors: $n = 3$). The L2 learners of English were Koreans living in Montreal, Canada. To ensure that the student participants had knowledge of the target vowels but difficulty perceiving them accurately, only intermediate learners of English were recruited, and their perceptual accuracy was measured during the pretest.

The L1 English speakers spoke general American English, and the L1 English listeners were either monolingual speakers of English or bilingual speakers of English and French residing in Montreal, Canada. The three ESL instructors all met the criteria described for the L1 English listeners group and had ESL teaching experience. Neither of the authors participated in this study as a L2 student, L1 English speaker, L1 English listener, or instructor.

Procedure

Table 1 summarizes the design of the study, which consisted of four steps: a pretest (administered to the L2 participants and also to the L1 English listeners along with the generalizability test to serve as

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Table 1. Design of the study

Steps	Procedures	Participants
Pretest (baseline test for the L1 English listeners)	Forced-choice identification tasks • For the L2 learners, 240 trials • For the English L1 listeners, 336 trials (240 trials + generalizability words, 96 trials)	(1) L1 English listeners (n = 16) (2) Instruction-only group (n = 16) (3) Instruction + CF group (n = 16)
Instruction (five 1-hr lessons)	(1) Pronunciation-focused instruction • Day 1: What is /i/-/ɪ/? Why is it important? • Day 2: /i/-/ɪ/ with /b/-/p/ • Day 3: /i/-/ɪ/ with /l/-/ɹ/ • Day 4: /i/-/ɪ/ with /ʃ/-/tʃ/ • Day 5: /i/-/ɪ/ with /f/-/v/ (2) Awareness tasks • Pick-a-card games • Bingo games • Fill-in-the-blank exercises (3) Corrective feedback • Two types of CF only for the instruction + CF group during the awareness tasks	(1) ESL instructors (n = 3) (2) Instruction-only group (n = 16) (3) Instruction + CF group (n = 16)
Immediate posttest	Forced-choice identification tasks • 336 trials (240 trials + generalizability words, 96 trials)	(1) Instruction-only group (n = 16) (2) Instruction + CF group (n = 16)
Delayed posttest (2 weeks later)	Forced-choice identification tasks • 240 trials	(1) Instruction-only group (n = 16) (2) Instruction + CF group (n = 16)

Note. L1 = first language; L2 = second language; CF = corrective feedback; ESL = English as a second language.

baseline data), the instructional treatment, an immediate posttest, and a delayed posttest. First, before the instructional sessions, 32 Korean learners of English and 16 native speakers of English took the pretest and the baseline test, respectively, in a research laboratory. Taking into account the L2 learners' age, gender, length of residence (LOR), and percent correct identification scores yielded by the pretest,

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

we assigned the 32 nonnative speakers to either the instruction-only group or the instruction + CF group (see Table 2). The groups were matched so that there were no significant differences on any of these variables.

The current study was based on a quasi-experimental design implemented in simulated classrooms rather than in intact classes. A total of four classes participated (i.e., two classes per group), each with eight learners. The instructional sessions consisted of five 1-hr pronunciation lessons, which were conducted for five consecutive days (i.e., 1 hr per day).

During every class, the respective instructional treatment was provided to both groups, whereas CF was given only to the instruction + CF group during the instructional tasks. All lessons were video recorded to confirm the consistency of the implementation of the two treatments. Finally, after the five 1-hr lessons, the learners took an immediate posttest. To determine whether the effects of the instruction and CF were retained after the instruction, the participants also took part in a delayed posttest 2 weeks later.

Instructional Treatment

To encourage learners to focus on the target vowels, the five 1-hr lessons were based on form-focused instruction. With its goal of drawing learners' attention to target forms that would otherwise not be noticed, form-focused instruction is known to have acquisitional value as an effective teaching technique (e.g., Loewen, 2014; Lyster, 2004b; Spada, 1997). The form-focused instruction implemented in this study included a combination of explicit instruction, input enhancement, and awareness tasks. First, in the case of explicit instruction, following Saito (2013), the instructors provided hyperarticulated pronunciation of the target contrast along with metalinguistic information, which entailed an explanation of why the contrast is important in English and why it is difficult for Korean learners of English to acquire. Second, typographically

Table 2. Participant information by group

Variable	Instruction-only (n = 16)	Instruction + CF (n = 16)
Age (years)	28.5 (*SD* = 5.9)	27.9 (*SD* = 5.3)
Male:female ratio (%)	25:75	25:75
LOR (months)	12.6 (*SD* = 13.7)	14.0 (*SD* = 14.0)
Total pretest scores (%)	58.2 (*SD* = 11.6)	58.9 (*SD* = 11.2)

Note. LOR = length of residence; CF = corrective feedback.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

enhanced input (Sharwood Smith, 1993) was employed to emphasize the target vowels by underlining and highlighting them in bold in the instructional materials. Third, awareness tasks were employed, following Lyster (2004a, 2007), to encourage learners to contrast the target vowels in contexts of games or exercises that provided opportunities for students to categorize the target vowels accurately and to consolidate their awareness of the contrasts.

Corrective feedback was provided to the instruction + CF group during the following three awareness tasks; learners participated in these tasks in each lesson, and they were used as a means to induce the learners to focus on /i/ and /ɪ/. The three awareness tasks were as follows:

- Pick-a-card games: Learners were given cards on which target words were written on either side. For example, *sheep* was written on one side, and *ship* on the other. The instructor asked each learner to pick up the card and show the side that corresponded with the word that the instructor had called out. When an individual learner failed to show the right side, immediate feedback was provided.
- Bingo games: Learners were asked to fill out a bingo card (5 × 5 square boxes) with daily target words. When the instructor called out a target word, learners were asked to mark it on their card, after which the instructor looked at their bingo cards one by one. In the case of wrong words being selected, the instructor provided immediate feedback to individuals. After verifying all the bingo cards, the instructor called the next word, and so on. Once a student completed a horizontal, vertical, or diagonal line, he or she called out "bingo" to win the game.
- Fill-in-the-blank exercises: This was a semidictation task in which learners were asked to fill in the target vowels. The instructor called out a word, and the learners wrote down the relevant sounds. As in the bingo games, the instructor checked every single worksheet and provided immediate feedback individually to learners if they were wrong. Once all of the worksheets were checked, the instructor called out the next word, and the exercise continued as such.

The instruction-only group participated in the same tasks, but the instructors provided no CF and no clues that might have lead students to believe that they were right or wrong. For instance, on the pick-a-card games, the instructor said, "the next word is . . ." regardless of whether they showed the right or wrong card. As for the bingo games and the fill-in-the-blank exercises, the instructor kept providing the next words without checking their answers and confirming whether they were right or wrong.

Focusing on the /i/-/ɪ/ contrast, the five 1-hr pronunciation lessons drew participants' attention to the target contrast within various consonantal environments. For instance, during Day 4, they practiced

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

the contrast between /ʃ/ and /tʃ/, such as in *sheep-ship* and *cheap-chip*. Regardless of the daily focus, all of the target words (see Table 3) were included in the awareness tasks, variations of which students completed in each of the five 1-hr lessons; at least approximately 10–15 minutes were allocated for each awareness task. Finally, the L2 learner participants were equally exposed to the three instructors to control for the influence of speaker variability in the instructional treatment.

Feedback Sequences

For the instruction + CF group, when perceptual errors occurred during the awareness tasks, feedback was immediately given to learners individually. It is important to note that the tasks were designed to improve the L2 learners' perceptual accuracy, so CF was given only on perceptual errors, not on production errors. In this regard, it has been suggested that self-repair following feedback on production errors is beneficial for language development because it engages learners in retrieval processes that stimulate the development of connections in memory (e.g., Lyster, 2004a). We predicted that self-repair following perceptual errors would have similar beneficial effects.

If learners responded in the right way, the instructor provided a confirmation with falling intonation such as "right, that is [correct word]," "perfect," "good," or "correct." If learners demonstrated an error in perception, the instructor immediately provided CF by uttering verbatim the word that was written or selected incorrectly by the learner with rising intonation and without any further explanation (referred to here as CF1). If self-repair took place, instructor confirmation and reinforcement followed, as illustrated by the excerpt in (1) from the pick-a-card game.

(1) Instructor: *Could you show me "fill"?*
Student: [Shows the instructor the card "feel"]
Instructor: *Feel?* → CF1
Student: *Umm.* [Shows the other side of the card, "fill"] *This?* → self-repair
Instructor: *Right, that is "fill"* [with falling intonation]. → instructor confirmation
Student: [Looks at the card "fill" again] *OK.*
Instructor: *Try again. Show me "fill."* → reinforcement

If the student did not self-repair in response to CF1, then the following CF, corrective feedback 2 (CF2), was provided: "I said [correct word],

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

but you picked up/wrote [incorrect word]. Right, then?" Corrective feedback 2 is tantamount to explicit correction, which conveys more explanation and contrastive information than CF1; the instructor produced both the target and the nontarget sounds as a means of conveying the message "your choice is not what I said." In addition, right after further explanation, the instructor provided an opportunity to self-correct by saying "then?" If the learners responded to this request in the right way, the instructor provided confirmation and reinforcement as in the CF1 sequence. The excerpt in (2) illustrates the use of CF2 during the fill-in-the-blank exercise.

(2) Instructor: *Next, "bit."*
Student: [Writes "beat" on the worksheet]
Instructor: *"beat?"* → CF1
Student: *Yes . . . Yes . . . Yes . . . ?* → no self-repair
Instructor: *I said "bit" but you wrote "beat." You see?* → CF2
Student: *Say again please.*
Instructor: *I said "bit" but you wrote "beat." Right, then?*
Student: *Aha, you said "bit." OK.* [Writes "bit" on the worksheet] *This?* → self-repair
Instructor: *Correct. Could you write "beat" next to it?* → instructor confirmation and reinforcement

Prior to the treatment, the instructors had been trained to either systematically provide CF or not. The video recordings, comprising 10 hr of the instruction-only treatment (two classes with five 1-hr lessons) and 10 hr of the instruction + CF treatment (also two classes with five 1-hr lessons), confirmed that no CF was provided on perceptual errors concerning the target form in the instruction-only classes, whereas CF was consistently provided in the instruction + CF classes. Specifically, approximately 450 CF turns were identified in the instruction + CF classes: 56% of these involved only CF1, after which immediate self-repair occurred. The remaining CF turns (i.e., 44%) were composed of double moves of CF1 and CF2; that is, as planned, if self-repair was not produced after CF1, the instructors provided the learner with CF2 to elicit self-repair. The need for instructors to employ CF2 proved to be task dependent; that is, during the fill-in-the-blank exercises, 85% of the CF turns consisted of a double move comprising both CF1 and CF2, whereas double moves accounted for only 30% and 22% of CF turns during the pick-a-card and bingo games, respectively.

After the feedback interaction (either CF1 only or double moves), self-repair occurred at a rate of almost 100%. In the rare instances when learners did not respond to CF2, topic continuation (e.g., the next word or another task)—rather than further provision of CF—ensued to move the lesson forward.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Testing Materials

A computerized test was designed specifically for the purpose of this study to assess the treatment effects. Because a series of studies confirmed that perceptual accuracy is influenced by speaker variability (Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan et al., 1991; Y. Wang, Spence, Jongman, & Sereno, 1999), each of the target words and distracters occurring in the testing materials was recorded by one male and one female general American English speaker, and the same sound files were consistently used for all test sessions.

Twelve sets of /i/-/ɪ/ pairs were collected from the Longman Communication 3000 corpus (*Longman Dictionary of Contemporary English*, 2009, pp. 2044–2059), which is a list of the 3,000 most frequent words in both spoken and written English. Moreover, to prevent participants from noticing the linguistic targets of the experiment, 12 sets of distracters (e.g., half of them with only the onset changed; the rest with only the nucleus changed) were also included, as outlined in Table 3. Items A: (1)–(8) were used as training stimuli, whereas items A: (9)–(12) were used as unfamiliar stimuli to see if the L2 learners could transfer the phonetic-level knowledge to new words.

Two L1 English speakers were asked to read the stimuli twice. The target stimuli were digitalized at 44,100 Hz, and one out of two productions was selected by each speaker's counterpart (i.e., the male speaker selected one out of two productions recorded from the female speaker, and vice versa). Finally, the sound files were analyzed with Praat (Boersma & Weenink, 2013), and it was confirmed that the L1 English speakers' productions of /i/ and /ɪ/ fell within the average of general American English speakers' vowel formants (Yang, 1996).

Previous studies (e.g., Flege et al., 1997) revealed that Korean learners of English are highly likely to exploit durational information to categorize the target vowels. To investigate the extent to which native and nonnative speakers of English rely on durational and spectral information, each word (except for distracters) was synthesized with five different durations in equal increments using a Klatt (1980) synthesizer. Each word was therefore prepared as 12 different tokens: Speaker (male and female) × Duration (natural, 130 ms, 160 ms, 190 ms, 220 ms, and 250 ms).

Using the PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993), forced-choice identification tasks were designed in which a minimal pair of stimuli and a relevant sound were given. If they were unsure about answers, participants were told to make a guess. There was no predetermined interval time between tasks. Instead, participants needed to press the "next" key before moving on to another task, so

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Table 3. Target stimuli

No.	A: Target words	B: Distracters
(1)	*beat-bit*	*book-look*
(2)	*cheap-chip*	*cut-shut*
(3)	*feel-fill*	*deal-seal*
(4)	*heat-hit*	*fat-hat*
(5)	*lead-lid*	*gun-sun*
(6)	*scene-sin*	*pin-thin*
(7)	*seat-sit*	*bag-big*
(8)	*sheep-ship*	*chat-cheat*
(9)	*leave-live*	*mass-miss*
(10)	*peak-pick*	*net-nut*
(11)	*reach-rich*	*rob-rub*
(12)	*seek-sick*	*tap-tip*

they could take as much time as they needed. Again, distracters were embedded alongside the target stimuli.

In the pretest, immediate posttest, and delayed posttest, the Korean learners of English were asked to complete a total of 240 trials (target stimuli: 16 words × 12 tokens = 192; distracters: 24 words × 2 tokens = 48). The immediate posttest consisted of the pretesting trials but in a different order and also included eight generalizability words (i.e., 96 trials) that had not appeared during the pretest and instruction; these unfamiliar trials were analyzed separately as a generalizability test. The delayed posttest did not include these unfamiliar trials and instead contained only the pretesting trials but in a different order. The 16 L1 English listeners took a baseline test composed of not only the 240 trials but also the 96 unfamiliar trials.

RESULTS

The scores yielded by this study are based on percentages of correct responses and were analyzed statistically, with an alpha level of .05. Two analyses—(a) a comparison of pretest, immediate posttest, and delayed posttest scores over time and (b) the generalizability test—were conducted separately using MANOVA and discriminant function analysis (DFA). In addition, to report the effect sizes of the instructional treatments, Cohen's *d* (Cohen, 1988) was also calculated and was classified as small ($0.40 \leq d < 0.70$), medium ($0.70 \leq d < 1.00$), or large ($1.00 \leq d$) for between-group contrasts but as small ($0.50 \leq d < 1.10$), medium ($1.10 \leq d < 1.60$), or large ($1.60 \leq d$) for within-group contrasts (Plonsky & Oswald, 2014). Scores of the L1 English listeners were used as a baseline for the analyses.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

The means, standard deviations, and 95% confidence intervals for the pretest, immediate posttest, and delayed posttest by group and by sound type appear in Table 4 and in Table 5 for the generalizability test (the baseline data were provided by the L1 English listeners). As for the pretest, immediate posttest, and delayed posttest, boxplots are displayed in Figure 1 for the instruction + CF group and in Figure 2 for the instruction-only group. Regarding the generalizability test, boxplots are illustrated in Figure 5 for both groups.

Comparison of Pretest, Immediate Posttest, and Delayed Posttest Scores over Time

To assess the effects of instruction and CF on the target vowels, a repeated-measures MANOVA was conducted with four dependent variables (/i/-natural, /i/-synthesized, /ɪ/-natural, and /ɪ/-synthesized), one within-group independent variable (time: pretest, immediate posttest, and delayed posttest), and one between-group independent variable (group: instruction-only and instruction + CF). With multiple dependent variables, a MANOVA is known to have greater statistical power to detect effects than an ANOVA because it considers the correlation between dependent variables and a combination effect of dependent variables (Huberty & Morris, 1989). In addition, by avoiding several sets of univariate analyses, it also controls for Type I error.

Before conducting the analysis, statistical assumptions—such as Shapiro-Wilk's 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—were confirmed. Most important, the result of Box's M test was not significant, $p = .305$. As a test statistic, Pillai's Trace was selected because it is the most powerful option when the samples are of equal size between groups.

The MANOVA revealed a significant effect for group, V = .43, $F(4, 27) = 5.12$, $p = .003$, and a significant within-group effect for time, V = .79, $F(8, 23) = 10.51$, $p < .001$. The analysis also identified a significant Group × Time interaction, V = .47, $F(8, 23) = 2.54$, $p = .038$. Accordingly, we conducted follow-up analyses of the interaction effect.

The default approach to follow up a significant MANOVA is to explore individual univariate ANOVAs for each dependent variable. However, according to Field (2013), this approach weakens the linear combination of the dependent variables verified by the MANOVA. Conducting several univariate analyses also increases Type I error. Accordingly, DFA was chosen to delve into the interaction effect. Discriminant function analysis is the mathematical inverse of MANOVA; that is, it suggests

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Table 4. Means, standard deviations, and 95% confidence intervals of the pretest, immediate posttest, and delayed posttest scores by group and sound type

Sound type	L1 listeners (n = 16)	Instruction-only (n = 16)			Instruction + CF (n = 16)		
	Baseline test	Pretest	Immediate posttest	Delayed posttest	Pretest	Immediate posttest	Delayed posttest
/i/-natural	99.8 (SD = 0.8) CI [99.4, 100]	68.4 (SD = 14.7) CI [60.5, 76.2]	85.2 (SD = 13.3) CI [78.1, 92.2]	81.3 (SD = 11.3) CI [75.3, 87.3]	72.7 (SD = 15.1) CI [64.6, 80.7]	84.1 (SD = 12.9) CI [77.2, 90.9]	91.4 (SD = 10.7) CI [85.7, 97.1]
/i/-synthesized	99.4 (SD = 1.4) CI [98.6, 100]	69.6 (SD = 16.6) CI [60.8, 78.4]	82.7 (SD = 14.0) CI [75.2, 90.2]	80.6 (SD = 5.16) CI [77.8, 83.4]	70.7 (SD = 13.6) CI [63.4, 78.0]	84.4 (SD = 11.5) CI [78.3, 90.5]	83.0 (SD = 14.2) CI [75.4, 90.5]
/ɪ/-natural	99.8 (SD = 0.9) CI [99.3, 100]	57.8 (SD = 15.7) CI [49.4, 66.2]	52.7 (SD = 23.4) CI [40.3, 65.2]	57.8 (SD = 20.9) CI [46.7, 68.9]	62.9 (SD = 16.4) CI [54.2, 71.6]	78.9 (SD = 12.3) CI [72.4, 85.4]	77.0 (SD = 15.9) CI [68.4, 85.4]
/ɪ/-synthesized	99.8 (SD = 0.7) CI [99.4, 100]	43.8 (SD = 19.2) CI [33.5, 54.0]	46.2 (SD = 21.1) CI [34.9, 57.4]	50.9 (SD = 18.7) CI [41.0, 60.9]	44.9 (SD = 11.2) CI [39.0, 50.9]	62.9 (SD = 20.4) CI [52.0, 73.8]	69.3 (SD = 14.9) CI [61.4, 77.2]

Note. Baseline data were provided by the L1 English listeners, abbreviated as L1 listeners. L1 = first language; CF = corrective feedback.

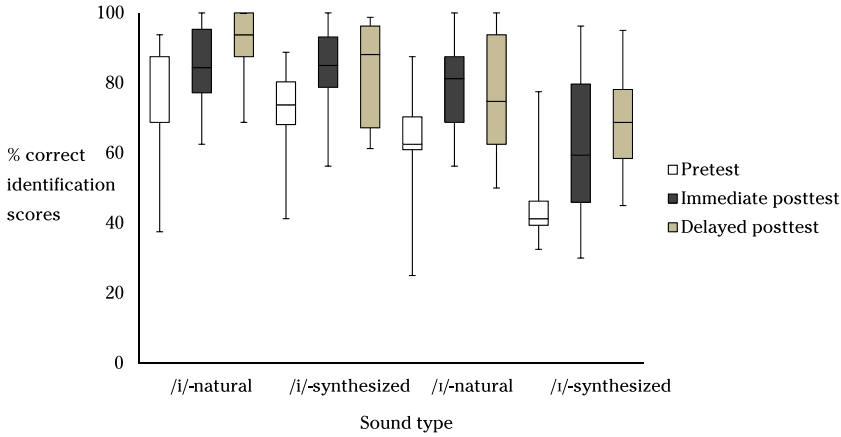


Figure 1. Boxplots of the pretest, immediate posttest, and delayed posttest scores of the instruction + CF group for each sound type.

how dependent variables in a MANOVA discriminate and predict a grouping variable.

A DFA was conducted to investigate the extent to which the dependent variables in the MANOVA discriminate and predict the group variable (i.e., the instruction-only or the instruction + CF group) at the three testing sessions. At the pretest, the function (canonical R^2 = .05) did not significantly differentiate the two groups, Λ = .95, $\chi^2(4)$ = 1.55, p = .818. In a similar vein, the effect sizes across all of the sound types at the pretest were small: /i/-natural (d = 0.30), /i/-synthesized (d = 0.08), /ɪ/-natural (d = 0.33), and /ɪ/-synthesized (d = 0.07).

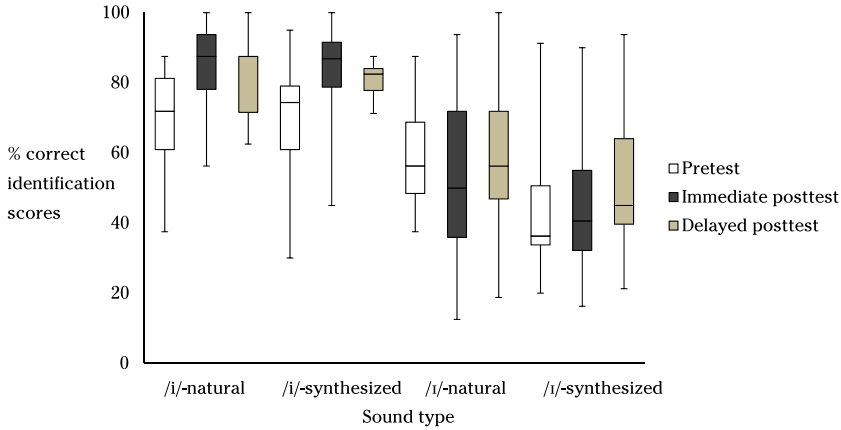


Figure 2. Boxplots of the pretest, immediate posttest, and delayed posttest scores of the instruction-only group for each sound type.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

However, at the immediate posttest, the function (canonical R^2 = .36) differentiated the two groups significantly, Λ = .64, $\chi^2(4)$ = 12.71, p = .013. The structure matrix found two important variables that differentiated the two groups, /ɪ/-natural (r = .96) and /ɪ/-synthesized (r = .55); however, /i/-synthesized (r = .09) and /i/-natural (r = –.06) were less important variables in differentiating the two groups. The values of the group centroids also confirmed that the function discriminated the instruction-only (–.73) and instruction + CF (.73) groups at the immediate posttest. The effect sizes also reflected the results of the DFA at the immediate posttest: a large effect size for /ɪ/-natural (d = 1.45) and a medium effect size for /ɪ/-synthesized (d = 0.83) but small effect sizes for /i/-synthesized (d = 0.14) and /i/-natural (d = 0.09).

In addition, the analysis indicated that the function (canonical R^2 = .35) significantly distinguished the two groups at the delayed posttest, Λ = .65, $\chi^2(4)$ = 12.03, p = .017. A detailed analysis of the structure matrix showed three significant predictors at the delayed posttest, /ɪ/-synthesized (r = .77), /ɪ/-natural (r = .73), and /i/-natural (r = .65), whereas /i/-synthesized was not significant (r = .16). As in the immediate posttest, the value of the group centroids also confirmed that the function differentiated the instruction-only (–.71) and instruction + CF (.71) groups at the delayed posttest. Effect sizes were also in the same vein; that is, large between-group effect sizes were found for /ɪ/-synthesized (d = 1.12) and /ɪ/-natural (d = 1.06), a medium between-group effect size was found for /i/-natural (d = 0.95), and a small effect size was found for /i/-synthesized (d = 0.23).

Another DFA was conducted to examine whether the dependent variables in the MANOVA discriminate and predict the time variable (i.e., pretest, immediate posttest, or delayed posttest) within the instruction-only and the instruction + CF groups, respectively. On the one hand, for the instruction-only group, the analysis found two discriminant functions, the first of which accounted for 95.9% of the variance with canonical R^2 = .28 and the second of which explained 4.1% with canonical R^2 = .02. The combination of these functions did not significantly differentiate the time variable, Λ = .71, $\chi^2(8)$ = 15.01, p = .059. The second function did not distinguish the time variable either, Λ = .98, $\chi^2(3)$ = 0.71, p = .870. Similarly, the effect sizes were calculated between the pretest and the immediate posttest and between the pretest and the delayed posttest. All of the effect sizes, except for /i/-natural between the pretest and the immediate posttest, were small: /i/-natural ($d_{\text{Pre.–Imm.}}$ = 1.19; $d_{\text{Pre.–Del.}}$ = 0.98), /i/-synthesized ($d_{\text{Pre.–Imm.}}$ = 0.85; $d_{\text{Pre.–Del.}}$ = 0.89), /ɪ/-natural ($d_{\text{Pre.–Imm.}}$ = 0.26; $d_{\text{Pre.–Del.}}$ = 0.00), and /ɪ/-synthesized ($d_{\text{Pre.–Imm.}}$ = 0.12; $d_{\text{Pre.–Del.}}$ = 0.37).

On the other hand, for the instruction + CF group, the first function explained 86.7% of the variance with canonical R^2 = .38, whereas the second accounted for 13.3% with canonical R^2 = .09. The combination

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

of the two differentiated the time points, $\Lambda = .57$, $\chi^2(8) = 24.52$, $p = .002$, whereas the second function on its own—decoupled from the combination—did not significantly differentiate the time points within the instruction + CF group, $\Lambda = .92$, $\chi^2(3) = 3.88$, $p = .275$. Specifically, with respect to the first function, the structure matrix revealed /ɪ/-synthesized ($r = .86$) and /i/-natural ($r = .78$) as significant predictors but /ɪ/-natural ($r = .58$) and /i/-synthesized ($r = .57$) as less powerful predictors. Regarding the second function, the structure matrix also found two significant predictors, /ɪ/-natural ($r = .64$) and /i/-synthesized ($r = .60$), whereas /i/-natural ($r = -.21$) and /ɪ/-synthesized ($r = .08$) were not significant. The values of group centroids confirmed that the first function differentiated the pretest (−1.02) from the immediate posttest (.24) and the delayed posttest (.78), and the second function distinguished the immediate posttest (.41) from the pretest (−.12) and the delayed posttest (−.29). Regarding within-group contrasts, effect sizes ranged from small to large in favor of the instruction + CF group: /i/-natural ($d_{\text{Pre.–Imm.}} = 0.81$; $d_{\text{Pre.–Del.}} = 1.43$), /i/-synthesized ($d_{\text{Pre.–Imm.}} = 1.09$; $d_{\text{Pre.–Del.}} = 0.88$), /ɪ/-natural ($d_{\text{Pre.–Imm.}} = 1.10$; $d_{\text{Pre.–Del.}} = 0.87$), and /ɪ/-synthesized ($d_{\text{Pre.–Imm.}} = 1.09$; $d_{\text{Pre.–Del.}} = 1.85$).

In sum, the repeated-measures MANOVA revealed significant main effects for group and time as well as a significant Group × Time interaction, following which DFAs were conducted to tease apart the interaction effect. The groups were indistinguishable at the pretest with respect to the four dependent variables, whereas the instruction + CF group outperformed the instruction-only group at the immediate posttest, owing to two significant predictors: /ɪ/-natural and /ɪ/-synthesized. Similarly, at the delayed posttest, the groups proved to be distinct as a result of three significant predictors: /ɪ/-synthesized, /ɪ/-natural, and /i/-natural. In the same vein, Cohen's *d* scores revealed that there were medium or large between-group effect sizes for the significant predictors, whereas small effect sizes were identified for the poor predictors.

For the instruction-only group, neither function significantly differentiated the pretest, the immediate posttest, and the delayed posttest scores, which, in contrast, were differentiated for the instruction + CF group. The first function, which had /ɪ/-synthesized and /i/-natural as the significant predictors, revealed significantly higher scores for the immediate and delayed posttests than for the pretest. The second function, which showed /ɪ/-natural and /i/-synthesized as the significant predictors, exhibited higher scores for the immediate posttest than for the pretest and the delayed posttest. All of the effect sizes, except for /i/-natural between the pretest and the immediate posttest, were small for the instruction-only group, whereas the effect sizes ranged from small to large for the instruction + CF group.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

As expected, the native speakers of English showed almost perfect perceptual accuracy (see Table 4) on all sound types. Levene's test for equality of variances turned out to be insignificant, $p = .102$; a one-way ANOVA found that there was no statistically significant difference between the sound types, $F(3, 60) = 0.75$, $p = .529$.

Across the within-group and the between-group analyses, the target phoneme /ɪ/ was a significant predictor for the group and time variables. This result also corresponded with the analyses of the synthesized stimuli with five different durations. Figures 3 and 4 illustrate the L2 learners' perceptual accuracy for each synthesized vowel across the five different durations at the pretest, the immediate posttest, and the delayed posttest. As expected, the responses of the Korean learners of English varied according to durational information. Both groups showed similar patterns: Concerning the /i/-synthesized sounds, the longer the vowel was, the higher the perceptual accuracy was. Conversely, regarding the /ɪ/-synthesized sounds, the shorter the vowel was, the higher the perceptual accuracy was. Not surprisingly, regardless of the durational manipulation of the testing words, the L1 English listeners showed almost perfect accuracy (more than 98.4% for the /i/-synthesized sounds and more than 99.5% for the /ɪ/-synthesized sounds) at 130 ms, 160 ms, 190 ms, 220 ms, and 250 ms.

As Figures 3 and 4 indicate, the Korean learners of English were better overall at the /i/-synthesized sounds than the /ɪ/-synthesized sounds. Despite the fact that both groups still relied on durational information after the instructional treatment, the most drastic difference between the two groups resulted from the /ɪ/-synthesized sounds. For the instruction-only group, there were no notable changes between the pretest, the immediate posttest, and the delayed posttest. However, for the instruction + CF group, the perceptual accuracy at the immediate posttest and the delayed posttest was generally higher than the accuracy at the pretest.

Generalizability Test

To investigate the extent to which the L2 learners perceived the target vowels in words that had not been included in the instructional sessions, the percent correct identification scores of the generalizability trials were compared (see Table 5 and Figure 5) between the instruction-only and instruction + CF groups using a one-way MANOVA with four dependent variables (/i/-natural, /i/-synthesized, /ɪ/-natural, and /ɪ/-synthesized) and one between-group independent variable (group: instruction-only and instruction + CF).

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

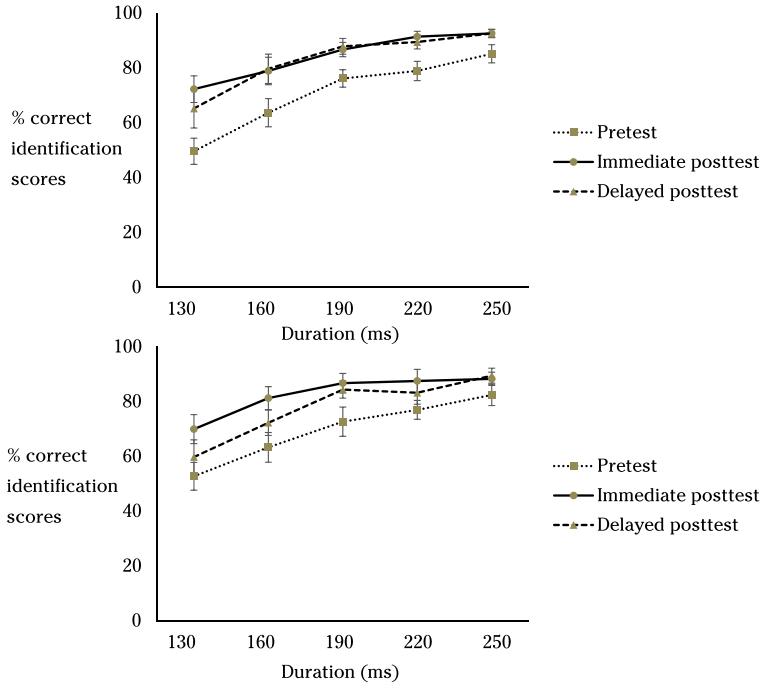


Figure 3. The use of durational information and effects of instruction (and CF) on the /i/-synthesized sounds (above: the instruction + CF group; below: the instruction-only group).

All assumptions for this statistical analysis were met, including the Shapiro-Wilk test for normality of distribution, a positive Pearson correlation between the dependent variables, and Levene's test for homogeneity of between-group variances. Box's M test was significant, $p = .003$, but the assumption regarding Box's M test is considered fulfilled because the p value was higher than .001. Moreover, when the sample sizes are equal, Box's M test is generally disregarded (for the statistical rationale, see Field, 2013).

Using Pillai's Trace, results of the MANOVA revealed that there was a significant effect for group on the four dependent variables, $V = .36$, $F(4, 27) = 3.75$, $p = .015$. To follow up on the significant main effect, a DFA was also conducted, and one discriminant function was found (canonical $R^2 = .36$), which differentiated the two groups significantly, $\Lambda = .64$, $\chi^2(4) = 12.38$, $p = .015$. Specifically, the structure matrix identified all of the dependent variables in the MANOVA as significant predictors: /ɪ/-natural ($r = .92$), /i/-synthesized ($r = .61$), /ɪ/-synthesized ($r = .53$), and /i/-natural ($r = .32$). The values of the group

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

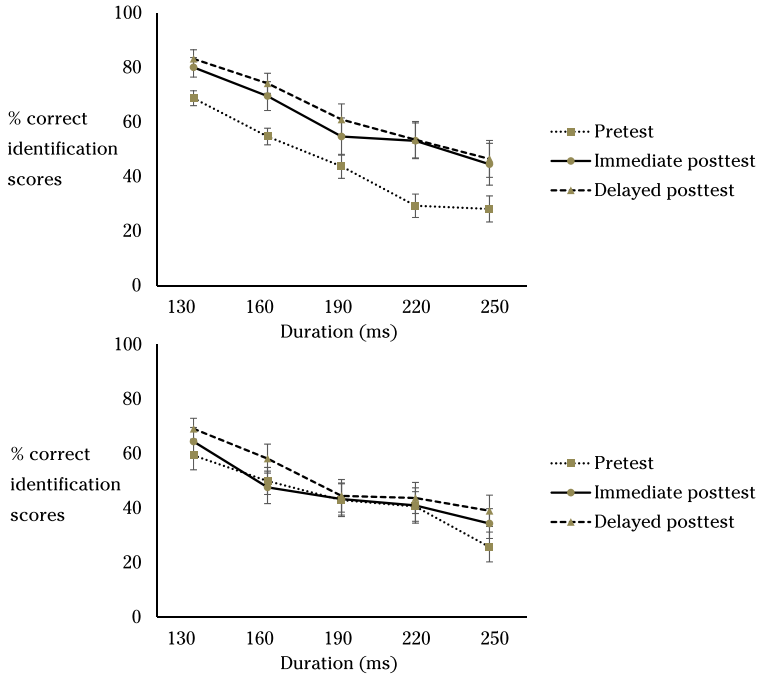


Figure 4. The use of durational information and effects of instruction (and CF) on the /ɪ/-synthesized sounds (above: the instruction + CF group; below: the instruction-only group).

centroids also verified that the function significantly distinguished the instruction-only group (–.72) and the instruction + CF (.72) group at the generalizability test.

On the basis of the descriptive statistics in Table 5, Cohen's *d* scores were calculated and revealed the following: a small effect for /i/-natural (d = 0.47), a medium effect for /i/-synthesized (d = 0.90) and /ɪ/-synthesized (d = 0.79), and a large effect for /ɪ/-natural (d = 1.37). As also seen in Table 5, the native speakers of English again showed more than 99.4% accuracy with no significant differences across all sound types: Because Levene's test was significant, p = .044, the F value was adjusted by Welch's F statistics, Welch $F(3, 32.4)$ = 0.59, p = .629.

In sum, the instruction + CF group performed better overall than the instruction-only group on the generalizability test, with all dependent variables in the one-way MANOVA turning out to be significant predictors distinguishing the two groups. Effect sizes ranging from small to large were identified in the between-group contrasts.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Table 5. Means, standard deviations, and 95% confidence intervals of the generalizability test scores by group and sound type

Sound type	L1 listeners (n = 16)	Instruction-only (n = 16)	Instruction + CF (n = 16)
/i/-natural	99.9	68.1	75.8
	(SD = 0.5)	(SD = 14.2)	(SD = 19.1)
	CI [99.6, 100]	CI [60.6, 75.7]	CI [65.6, 85.9]
/i/-synthesized	99.4	62.5	79.1
	(SD = 1.3)	(SD = 17.5)	(SD = 20.3)
	CI [98.7, 100]	CI [53.2, 71.8]	CI [68.2, 89.9]
/ɪ/-natural	99.8	51.4	76.1
	(SD = 0.6)	(SD = 18.0)	(SD = 19.2)
	CI [99.5, 100]	CI [41.8, 61.0]	CI [65.8, 86.4]
/ɪ/-synthesized	99.7	37.7	53.4
	(SD = 0.7)	(SD = 9.6)	(SD = 27.5)
	CI [99.4, 100]	CI [32.5, 42.8]	CI [38.7, 68.1]

Note. Baseline data were provided by the L1 English listeners, abbreviated as L1 listeners. L1 = first language; CF = corrective feedback.

DISCUSSION

The Effects of CF on Instructed L2 Speech Perception

The most noteworthy finding of this study is that the instruction + CF group significantly outperformed the instruction-only group as a result of the instructional sessions specifically with respect to the use of

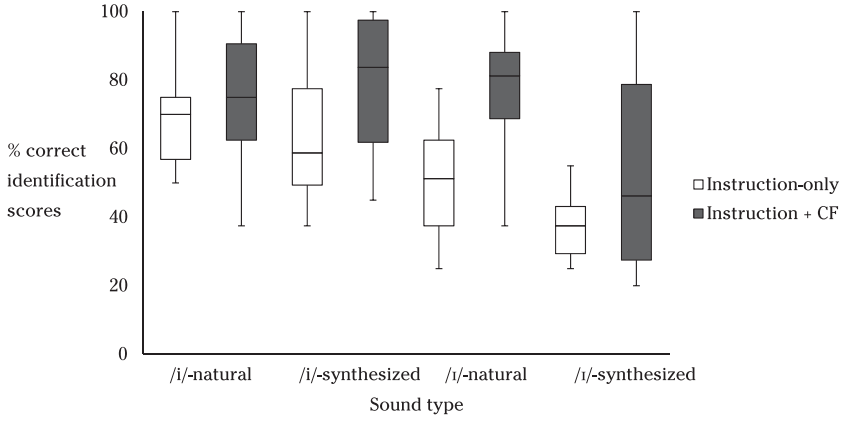


Figure 5. Boxplots of the generalizability test scores of the instruction-only and instruction + CF groups for each sound type.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

/ɪ/-natural and /ɪ/-synthesized sounds. Whereas the phoneme /i/ belongs to the Korean vowel inventory, the phoneme /ɪ/ does not, in that the target contrast becomes allophonic rather than phonemic in the learners' L1. However, the contrast is crucially phonemic in English and thus leads to perceptual confusion in the case of Korean learners of English.

As in previous studies (e.g., Baker et al., 2002; Tsukada et al., 2005), the current study found that Korean learners of English showed higher perceptual accuracy when /i/ was provided as auditory input (71% in the pretest) than when /ɪ/ was provided (51% in the pretest). The PAM (Best, 1995) and PAM-L2 (Best & Tyler, 2007) predict that the two nonnative phonemes /i/ and /ɪ/ will be assimilated into the native phoneme /i/. Specifically, the models predict that Korean learners of English will show an assimilation pattern of either single category or category goodness. In the present study, a category-goodness assimilation pattern was found, given the participants' higher accuracy on the target phoneme /i/ than on /ɪ/. In addition, the NRV framework (Polka & Bohn, 2011) predicts that /i/ is the natural referent vowel because of its acoustic properties and perceptual salience. The model also predicts that the natural reference vowel becomes more prominent when L2 learners acquire a new vowel system.

The findings in the present study thus allow us to conclude that CF played a key role in helping the L2 learners to acquire the target phoneme /ɪ/. As a result of the instruction and CF inducing the Korean learners of English to be aware of the spectral contrast between /i/ and /ɪ/, they were, in turn, able to reduce the perceptual confusion. With the support of form-focused instruction that included explicit crosslinguistic information, input enhancement, awareness tasks, and CF, the L2 learners in the present study were encouraged to notice the existence of two different phonemes in the target language.

The instructional tasks provided the L2 learners with opportunities to reflect on and modify their linguistic knowledge. The awareness tasks requiring participants to differentiate /i/ and /ɪ/ provided them with ample opportunities to test their hypotheses about this problematic contrast. In addition, in the case of the instruction + CF group, CF on the learners' errors was immediately provided, such that their linguistic knowledge was challenged by negative evidence that pushed them to find the right representation. Such demands induced the learners to notice the gap (see Schmidt & Frota, 1986) between their current knowledge and more targetlike knowledge and to restructure their existing knowledge toward targetlike representations.

When the instructor asked learners to pick a card and show the side with a word containing /ɪ/, they were confronted with a decision (i.e., /i/ or /ɪ/). Owing to the absence of the target phoneme /ɪ/ in their L1, the learners were predisposed to making a number of perceptual errors. If those in the instruction + CF group were wrong, CF incited them to notice

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

that their current linguistic representation was out of line with the target language and pushed them to restructure their knowledge representation; accordingly, CF helped the learners to become aware of the spectral contrast between the two target phonemes.

Finally, with the instructors' confirmation, such as "right, that is [correct word]," "perfect," "good," or "correct," the learners were able to consolidate their linguistic knowledge. Through the procedure of reinforcement embedded in the feedback sequence of this study, the learners were again asked to categorize the target phonemes accurately and, thus, had another opportunity to confirm their linguistic knowledge.

Awareness Tasks and Types of CF

To create contexts of interaction between instructors and students and to allow for the provision of CF on perceptual errors, three types of awareness tasks were employed during the lessons. In terms of CF, two types of CF were provided on perceptual errors. Specifically, CF1 (a verbal repetition of the error) was delivered with rising intonation. Then, if no self-repair was elicited by CF1, CF2 (an explicit correction: "not X but Y") was provided.

Repetition is known as one type of prompt that signals an error and pushes learners to self-repair. According to Ellis (2006) and Loewen and Nabei (2007), repetition may be considered implicit because it has proven less likely to lead to self-repair. However, this study found that whether repetition resulted in self-repair depended on the nature of the task. During the fill-in-the-blank exercises, 85% of the feedback turns consisted of a double move comprising both CF1 and CF2, whereas double moves accounted for only 30% and 22% during the pick-a-card and bingo games, respectively. That is, during the pick-a-card and bingo games, repetition was explicit enough to induce the learners to notice their errors and to result in accurate self-repair; however, during the fill-in-the-blank exercises, the learners required more elaborated negative evidence beyond repetition to notice their errors, as we explain next.

Concerning the pick-a-card and bingo games, the target pairs (e.g., *sheep-ship*) were written on the cards. The games were recognition tasks requiring the learners to consider only the sounds of the target vowels. For example, once students inferred negative evidence from CF1, they were easily able to notice that the error was related to the target vowel and to find the other vowel of the minimal pair.

With respect to the fill-in-the-blank exercises, learners needed to write relevant vowel(s) in response to what instructors said. These exercises required two kinds of linguistic knowledge: (a) phonological

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

(/i/ or /ɪ/) and (b) orthographic. For example, when asked fill in the blank *S___t* on hearing "seat," the learners needed to determine whether it was /sit/ or /sɪt/ at the phonological level and then to write the relevant letters based on their orthographic representation (e.g., *seat*, *seet*). When given CF1 with rising intonation by the instructor, there was potential for the learners to be confused about whether the feedback addressed the phonological representation (i.e., /i/ or /ɪ/) or the orthographic representation (i.e., *ea* or *ee*), or both. Due to the lack of specific information in CF1 addressing errors in the fill-in-the-blank exercises, the learners may not have felt able to immediately self-repair, and so CF2 ensued with further explanation.

LIMITATIONS AND FUTURE DIRECTIONS

The current study has several limitations that restrict its generalizability. First, a sample size larger than 32 would, of course, strengthen the statistical analyses and findings. Second, the generalizability trials were measured only at the time of immediate posttesting. Including them as pretesting trials would make our speculation that the benefits of CF can extend to novel words not occurring in the instructional materials much less tentative. Third, our use of simulated classrooms in this study begs the question as to whether such intense instruction would be feasible in a regular classroom curriculum and whether the results would be similar.

These limitations can be seen as providing the impetus for further research in this area that would include larger samples, more robust measures of novel words, and real rather than simulated classrooms. Our goal was to explore the possibility of L2 speech perception training in a setting that approximated classroom conditions to a greater degree than laboratory settings are able to. The present study thus provides the basis on which future classroom studies could be built. Also of interest would be to further explore the finding that task type had a clear impact on the types of CF that were used by instructors; worthy of further pursuit in this regard would be the extent to which the different tasks employed in this study might differentially impact L2 learners' perception accuracy.

CONCLUSION

"No speaker of any language perceives acoustic reality; in each case, perception is altered in the service of language" (Kuhl, 2000, p. 11852). In the present study, that which is "in the service of language" could be thought of as the L2 learners' perceptual system, which is intrinsically

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

linked to L1 background. Specifically, it is argued that Korean learners of English have difficulty perceiving /i/ and /ɪ/ owing to expectations carried over from their L1. However, it is also important to stress that perception is alterable with appropriate language exposure. Specifically, with respect to L2 speech perception, considering that a great deal of EFL learning takes place in classroom settings, appropriate language exposure occurs mainly through instruction, teaching aids, and interaction between instructors and students. In this regard, the present study sheds light on the role of CF in L2 speech perception training in classroom settings. During the instructional tasks, L2 learners had opportunities to retrieve and test out their own linguistic knowledge. If it proved unacceptable in the target language, CF interaction provided them with opportunities to restructure their knowledge, ultimately leading to the right representation.

Previous research findings regarding laboratory-based perception training have emphasized the role of CF. However, what has so far received little attention is why and how CF helps L2 learners to perceive nonnative phonemes correctly during the training sessions. In this regard, the current study suggests a pivotal role for CF and its importance in speech perception training. That is, CF provides opportunities for L2 learners to (re)confirm their L2 linguistic knowledge and can lead to a restructuring of the interlanguage system and a consolidation of targetlike knowledge.

Many SLA studies have found effects for CF on L2 grammatical, lexical, and pragmatic targets and, recently, on L2 speech production (Lyster, Saito, & Sato, 2013). Given that there are no previous studies investigating the effects of feedback on L2 speech perception, this study is expected to expand horizons in regard to the roles attributed to CF.

In spite of the importance of speech perception, L2 pedagogical issues have been discussed predominately with respect to speech production. Yet Borden, Gerber, and Milsark (1983) suggested that perception abilities may be a prerequisite for successful L2 speech production. In pursuing the question of how to teach L2 pronunciation most effectively, the fruitful solution to this question may lie in the realm of L2 speech perception.

Received 5 November 2013
Accepted 17 October 2014
Final Version Received 31 October 2014

REFERENCES

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. 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.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues 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 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, the Netherlands: Benjamins.
Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.41) [Computer program]. Retrieved from http://www.praat.org/
Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Timonium, MD: York Press.
Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303–328.
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.
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.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.
Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers*, *25*, 257–271.
Ellis, R. (2006). Researching the effects of form-focussed instruction on L2 acquisition. *AILA Review*, *19*, 18–41.
Escudero, P. (2006). Second language phonology: The role of perception. In M. Pennington (Ed.), *Phonology in context* (pp. 109–134). New York, NY: Palgrave Macmillan.
Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.
Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London, UK: Sage.
Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.
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.
Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*, 495–522.
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.
Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*, 302–308.
Jamieson, D. G., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, *16*, 201–210.
Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.
Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA*, *97*, 11850–11857.
Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & 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.
Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (Eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech* (pp. 72–97). Amsterdam, the Netherlands: University of Amsterdam.
Leather, J. (1999). Second-language speech research: An introduction. In J. Leather (Ed.), *Phonological issues in language learning* (pp. 1–58). Oxford, UK: Blackwell.
Lee, H. (1993). Korean. *Journal of the International Phonetic Association*, *23*, 28–31.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

Lee, K. (2008). *The effect of perceptual training on the perception and production of English vowel contrasts by Korean speakers* (Unpublished master's thesis). Hanyang University, Seoul, South Korea.
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.
Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.
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.
Loewen, S. (2014). *Introduction to instructed second language acquisition*. New York, NY: Routledge.
Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 361–377). Oxford, UK: Oxford University Press.
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.
The Longman Dictionary of Contemporary English (5th ed.). (2009). New York, NY: Pearson Longman.
Lyster, R. (2004a). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.
Lyster, R. (2004b). Research on form-focused instruction in immersion classrooms: Implications for theory and practice. *Journal of French Language Studies*, *14*, 321–341.
Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam, the Netherlands: Benjamins.
Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*, 1–40.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.
Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*, 467–478.
Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues 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: Issues in cross-language research* (pp. 411–432). Timonium, MD: York Press.
Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*, 842–854.
Saito, K. (2013). 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.
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.
Saito, K., & Lyster, R. (2012b). Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly*, *46*, 387–398.
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*, 1–34.
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*

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press

to learn: Conversation in second language acquisition (pp. 237–326). Rowley, MA: Newbury House.
Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*, 40–72.
Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication*, *45*, 425–434.
Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*, 165–179.
Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *29*, 73–87.
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.
Thomson, R. I. (2007). *Modeling L1/L2 interactions in the perception and production of English vowels by Mandarin L1 speakers: A training study* (Unpublished doctoral dissertation). University of Alberta, Canada.
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.
Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production* (Unpublished doctoral dissertation). Simon Fraser University, British Columbia, Canada.
Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*, 539–552.
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.
Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658.
Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, *24*, 245–261.

https://doi.org/10.1017/S0272263115000194 Published online by Cambridge University Press