

## **Feedback on Local Intelligibility by Same-L1 Learners: Face-to-face or Computer-mediated?**

**Suzanne Marie Yonesaka**

Hokkai-Gakuen University

<suzanne@hgu.jp>

### **Abstract**

During spoken interaction with peers, ESL or ELF learners use peer feedback to negotiate and converge on mutually intelligible pronunciation. In contrast, for EFL learners who interact mainly with same-L1 peers, it is not clear whether the giving and receiving of peer feedback on intelligibility can facilitate uptake leading to improved pronunciation. However, it has been suggested that same-L1 learner pronunciation feedback requires ample time and a narrow focus, conditions that could be realized in a CMC environment. This study investigates the effect of modality—face-to-face (FtoF) or computer-mediated (CMC)—of peer feedback on the intelligibility of 83 Japanese university students' /l/ and /r/ in four target words. After discrimination and pronunciation training, treatment consisted of participants making peer intelligibility judgments in two modalities. The FtoF group pronounced and provided immediate feedback in pairs, while the CMC group recorded utterances that were randomly delivered to peers for asynchronous intelligibility judgements. The participants' pre-, post-, and delayed post-test recordings (956 utterances) were rated for intelligibility. For each target word, Generalized Estimating Equations examined the effect of group, initial discrimination ability, change in listening discrimination ability, and initial pronunciation confidence. The CMC group was significantly more likely to be intelligible for post-treatment singleton-/r/ and delayed post-treatment /l/-clusters. The results suggest that, depending on the intended target, intelligibility feedback from same-L1 peers in a CMC environment does not hinder and may promote pronunciation development.

### **Introduction**

In EFL contexts, learners' opportunities to speak the target language occur primarily with same-L1 peers during classroom-based interaction. Learners' utterances include a range of pronunciations—some closer to the target language, some more strongly influenced by the L1, and some unique to the speakers—that serve as listener input. During classroom communication, learners provide and receive feedback, some of it related to pronunciation, signaling whether utterances have been understood. Although such interaction is ubiquitous, it is not clear to what extent learners actually use pronunciation input and feedback from same-L1 peers to improve

their own pronunciation. Pronunciation research has expanded beyond the inner-circle/native-speaker paradigm, and research within the English as a Lingua Franca (ELF) paradigm has found that learners use pronunciation input and feedback from different-L1 learners to negotiate effective forms of pronunciation (Sicola, 2009; O’Neal, 2015). However, as pointed out by Walker (2001), the current reality is that most learners are interacting with same-L1 peers. When it comes to input and feedback from same-L1 learners, research provides relatively little guidance for pronunciation classroom practitioners. To better understand same-L1 peer input and feedback, it is helpful to disentangle peers’ roles as speakers (feedback receiver) and listeners (feedback giver). If uptake occurs, that is, if pronunciation improves, to what extent is it due to the learner listening carefully and noticing in order to give feedback? To what extent is it due to the feedback received? If the former, a computer-mediated environment that allows for repeated listening would enhance the role of the feedback giver. If the latter, a face-to-face environment with real-time interaction could enhance affective aspects for the feedback receiver. This study, which is theoretically grounded in the Intelligibility Principle which “holds that the goal is intelligible speech, irrespective of how native-like it sounds” (Derwin & Munro, 2015, p. 6), contributes to the research base by examining the effect of same-L1 peer feedback on the intelligibility of Japanese university EFL learners under two conditions: face-to-face (FtoF) and in computer-mediated communication (CMC).

## Literature Review

### Peer feedback

Providing corrective feedback and support has been commonly considered to be one of the pronunciation teacher’s most important responsibilities (Morley, 1991; Szypra-Kozłowska, 2014). This is not surprising, as individual corrective feedback on pronunciation from the instructor has been found to be a significantly more powerful teaching tool than listening-only activities (Dlaska & Krekeler, 2013). Japanese learners also consider corrective feedback on pronunciation to be important. In a survey of 1228 Japanese university students, Timson, Grow, and Matsuoka (1999) found that about 60% desired corrective feedback on every pronunciation error and nearly 90% wanted this feedback to come from the instructor.

Despite this learner preference for teacher feedback, peer feedback is commonly used (Foote, Holtby, & Derwing, 2012). Without corrective feedback, pronunciation errors can fossilize; as an additional source of feedback besides the instructor, peer feedback on pronunciation can help prevent this. Providing individual pronunciation is time-consuming for instructors of relatively small classes (Baker, 2011; Baker & Burri, 2016), but even more so for the large classes found in many EFL contexts (Bahanshal, 2013), and peer feedback can reduce this heavy burden. Instructors may also use peer feedback for affective reasons. For learning in general, peer feedback can foster students’ motivation and sense of ownership because it involves them directly in the learning process (Topping, 1998). Pronunciation learners have also felt peer corrective feedback to be useful and motivating (Kim & Lee, 2015; Lord, 2008; Yoon & Lee, 2009).

One pedagogical reason for including peer feedback is that it supports self-assessment. Learners are notoriously poor at accurately self-assessing their pronunciation skills (Dlaska & Krekeler, 2013; Trofimovich, Isaacs, Kennedy, Saito & Crowther, 2016); however, peer feedback can cultivate the dispositions and skills needed for self-assessment (Chang, 2012; Roccamo, 2015; Topping, 1998). Another pedagogical reason for peer feedback is uptake—improved pronunciation by the feedback receiver. Kim and Lee (2015) analyzed the online peer

pronunciation of students who worked in groups to provide corrective peer feedback, finding that the students' later recordings did incorporate feedback from earlier recordings.

There is some evidence that, in line with what has been shown in writing assessment (Knoch, 2017), the feedback provider also receives learning benefits. Luo (2016) examined the effect of peer feedback on pronunciation development. The experimental group recorded written passages and posted them online, then worked in small groups so that each student received written pronunciation feedback from three students. The experimental group showed the largest pronunciation gains, with most of the learners attributing this to having listened to their peers' recordings. In Counselman's (2010) study, the control group recorded a dialogue and received instructor feedback. The experimental group did not record but listened to the control group's recordings and annotated the script for non-target-like pronunciation. Only the experimental group showed significant improvement, and Counselman concluded that providing fine-grained feedback to peers at a phonetic level allowed learners to notice the features effectively and efficiently. Both of these studies support the idea that intensive listening to peers' speech can lead to improvement in production, even for same-L1 peers.

However, studies also suggest that providing corrective pronunciation feedback is not easy. Learners find it difficult to pinpoint problematic areas (Luo, 2016), lack enough knowledge to provide accurate feedback (Lyster, Saito, & Sato, 2013), or provide feedback that is too indirect to be useful (Kim & Lee, 2015). For these reasons, rather than having peers search for pronunciation errors and provide suggestions or explanations, the present study asks them to indicate that an error has occurred by indicating whether or not they heard the intended target word of a minimal pair. Indicating errors is considered to be the least challenging level of corrective feedback (Ellis, 2009; Topping, 1998). Unambiguous information from peers as to whether or not the intended phoneme was perceived will alert learners that a pronunciation problem exists, the first step in pronunciation improvement. Language learners have difficulty identifying their own specific pronunciation problems (Derwing & Rossiter, 2002) and tend to believe their pronunciation is more accurate than it is (Lefkowitz & Hedgcock, 2002). When learners discover that a specific pronunciation feature is not intelligible—even to their same-L1 peers—they may be more likely to notice that feature and attempt modifications.

### **Intelligibility**

Intelligibility has been defined as “the extent to which the acoustic-phonetic content of the message is recognizable by a listener” (Field, 2005, p. 401) and as “the extent to which a listener actually understands an utterance” (Derwing & Munro, 2005, p. 385). The intelligibility principle is directly connected to EFL pronunciation teaching because, unlike a focus on accentedness with native-like pronunciation as a goal, it emphasizes listener comprehension even if the speech differs from that of the listener (Munro & Derwing, 2015a). Intelligibility has been measured impressionistically through scalar judgments on a continuum, most commonly with nine points (Isaacs, 2008), and it has also been measured objectively by having listeners complete cloze passages (Matsuura, 2007) or transcribe utterances (Derwing & Munro, 1997; Zielinski, 2008). This study uses Derwing and Munro's (2005) definition of intelligibility, operationalizing it as a forced-choice judgment of *local intelligibility* (Munro & Derwing, 2015b), that is, the intelligibility of a particular pronunciation feature (in this case, contrasting phonemes), in a highly-controlled feedback activity in which the hearer chooses which contrastive utterance was perceived.

This study's focus on local rather than global intelligibility is also intended to minimize the effects of the Interlanguage Speech Intelligibility Benefit (ISIB) (Bent & Bradlow, 2003), which

predicts that a foreign speaker will be as intelligible to a foreign listener as a native speaker if the foreign speaker and listener share the same L1. Studies both support and refute the ISIB. However, in a meta-analysis using linear modeling, Wang and van Heuven (2015) found a clear binary division: “when two interactants share the same native language, they enjoy the advantage of a shared phonology... [and when they do not,] their mutual intelligibility is poorer” (p. 10).

### **Modality**

The present study uses two modalities to explore which aspects of same-L1 peer feedback are most salient for learners. Face-to-face (FtoF) is synchronous and audio-visual; in this study, computer-mediated communication (CMC) is asynchronous and audio. Each offers different advantages for pronunciation feedback.

In face-to-face communication, the listener has access to both audio and visual input, termed the audio-visual benefit. For native-speaker listeners, visible gestures can provide some cues as to the place of articulation of vowels and the place and manner of articulation of consonants (Hazan et. al, 2006). Prosodic cues tend to be located in the upper part of the face, and experienced listeners direct more gazes toward the upper facial region when making stress and intonation decisions (Lansing & McConkie, 1999; Swerts & Kraemer, 2008). For language learners, the audio-visual benefit is likely to increase speech intelligibility and reduce the likelihood of pronunciation-related communication breakdowns during interaction (Loewen & Isbell, 2017). However, lower proficiency learners (Sueyoshi & Hardison, 2005) and Japanese learners, who direct their attention mainly to auditory cues (Hazan et. al, 2006), benefit less from the visual channel. Loewen and Isbell also note that although FtoF is the baseline for modality comparisons, its dual channels (audio and visual) can increase the cognitive load for learners to the point of overload.

Computer-mediated communication can be synchronous or asynchronous, voice-only or audio-visual. Bueno-Alastuey (2011, 2013) found that, although synchronous voice-only CMC helped learners notice and negotiate pronunciation during information gap exercises, same-L1 NNS dyads produced significantly less modified pronunciation output than different-L1 dyads. She recommends researchers to explore ways to increase learner noticing and modified output for same-L1 dyads; thus, the present study uses asynchronous voice-only CMC. Listening to recorded speech that can be repeated and stored as opposed to real-time speech provides pronunciation learners with the time and opportunity for critical listening to their own speech and to the speech of their peers, allowing learners to notice features that they otherwise would not (Correa & Grim, 2014; Gilakjani, Ahmadi, & Ahmadi, 2011). These improved perceptual skills are critical for pronunciation improvement. Flege’s Speech Learning Model (SLM) predicts that if learners are able to distinguish L2 from L1 sounds at a perceptual level, then production will eventually become more accurate (Derwing & Munro, 2015). Another rationale for recorded speech is that convergence—the gradual accommodation of speech features to those of the interlocutor—can occur during real-time communication tasks for same-L1 learners, leading to increased intelligibility but reinforcement of the L1 accent (Walker, 2005; Walker & Zoghbor, 2015).

### **Targeted phonemes**

According to Flege’s speech learning model (SLM), the perception and production of L2 segmentals are affected by two processes. Category dissimilation is when the learner perceives a difference between the existing L1 sound and the new L2 sound, and is able to form a new category. Category assimilation is when the speaker has difficulty forming a new category for the L2 sound because the L1 has a similar sound (Franklin & Stoel-Gammon, 2014). The present study focuses on the /l/-/ɹ/ contrast, a complicated example of category assimilation for Japanese

learners. Although the Japanese alveolar liquid /ɾ/ and the English liquids /l/ and /ɹ/ differ in place and manner of articulation, Japanese learners perceptually collapse the English liquids into the same category as the Japanese, leading to sound substitutions (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). Japanese and native-speaker pronunciation raters (Kashiwagi & Snyder, 2008), instructors (Saito, 2014) and learners (Ishikawa, 2015) identify the /l•/ɹ/ distinction as being particularly important for intelligible speech. In contrast, O’Neal (2013) found that the Japanese /ɾ/ can substitute for /l/ or /ɹ/ without inhibiting the negotiation of intelligibility during Japanese learner interaction with speakers of other first languages.

For adults, it may take several weeks of training with high variability stimuli for Japanese learners to be able to perceive (Lively, Logan, & Pisoni, 1993) and produce (Bradlow, Akahane-Yamada, Pisoni & Tohkura, 1999) the /l•/ɹ/ contrast, and there is some evidence (Sheldon & Strange, 1982) that even advanced Japanese learners who can pronounce these phonemes can continue to have difficulty perceiving them, and vice-versa (Hattori, 2010). Japanese university students have been found to continue over time to substitute the Japanese /ɾ/ for /l/ in singleton onsets, a substitution that contributes to their global foreign accent (Riney, Takada, & Ota, 2000). As for children, Akatsuka and Yamami (2016) found that long-term low-input English conversation training had no effect on Japanese elementary learners’ discrimination or intelligibility of initial singleton-/l/ or /ɹ/. In contrast, Aoyama et al. (2004) found that both child and adult Japanese learners showed more longitudinal improvement in production of /ɹ/ than for /l/, despite /l/ being easier to pronounce. Because /ɹ/ is perceived as highly dissimilar, it is easier for learners to form a new phonetic category.

The research questions are:

1. To what extent does modality of same-L1 peer pronunciation feedback (face-to-face or computer-mediated) affect intelligibility of the /l•/ɹ/ contrast?
2. To what extent does target variability (*lamp, ramp, glass, grass*) affect intelligibility?
3. To what extent do learner factors (initial pronunciation confidence, initial ability to discriminate /l/ and /ɹ/, and change in listening discrimination ability) affect intelligibility?

The hypothesis (one-tailed) was that the use of computer-mediated peer feedback would increase the proportion of intelligible utterances.

## Methodology

### Participants

Participants were 83 students (M=42, F=41) enrolled in two sections of a practical English phonetics course at a mid-level private Japanese university outside of Tokyo. Of these, 73 were first-year students; 71 majored in English language and culture and 12 majored in Japanese language and culture. No participants were hearing-impaired, all had Japanese as their L1, and all provided informed consent following TESOL guidelines. The 100 students originally enrolled in the course had been randomly divided into two equal groups, but attrition, lack of informed consent, or technology difficulties during testing resulted in unbalanced control and experimental groups (FtoF=38, CMC=45). TOEIC-RL scores were available for 65 of the participants; these ranged from 260 to 685, with a mean of 439, somewhat lower than the average score of 464 for first-year English majors nationwide in 2016 (IIBC, 2017). Boxplots showed that the TOEIC scores were normally distributed for both groups, and an independent samples t-test found no significant difference in mean scores, indicating that the two groups had comparable receptive skills, at approximately CEFR level A2.

To assess whether they adequately represented the larger population of Japanese university students, participants responded to demographic questions regarding their English use and to a survey (Otsuka & Ueda, 2011) on their previous pronunciation learning and present pronunciation confidence. Only 19% had been in an English-speaking country for up to a month and none more than that; 70% claimed they had little or no opportunity to engage in real-life communication outside the classroom.

As for high school pronunciation instruction or practice, reading aloud the main passage in the textbook was the most commonly cited activity (83%). Reading aloud is generally performed as a whole-group fluency production activity, possibly with generalized pronunciation feedback to the entire class. The second most-commonly cited was word stress (80%), an area that receives attention because some university entrance exams contain written vocabulary recognition items in which testees choose the word with the same stress pattern as the target word. English rhythm was the area in which fewest participants had received instruction (27%). Although 55% of participants claimed to have had at least one high school teacher with good pronunciation, only 39% felt that pronunciation had been emphasized. Only 14% of participants felt that their high school teachers had attempted to speak as little Japanese as possible. Although 41% of participants indicated that pronunciation testing had been conducted, the recording of pronunciation in class was extremely uncommon (8%). It is probable that pronunciation testing had involved not spoken production but written vocabulary recognition of stress or IPA. These responses, as well as responses to an open-ended question, suggest that participants had received pronunciation instruction or practice typical of non-elite high schools in Japan.

The participants also rated eight can-do statements about their present ability in various areas of pronunciation (Otsuka & Ueda, 2011). Overall, the participants had little confidence in their pronunciation ability. About one-third of them felt they could understand and read word stress, express emotions while reading, and understand rising and falling intonation, and less than one-fifth felt they could produce English sounds correctly or read with English rhythm. Overall, the participants can be judged to be representative of first-year students at non-elite universities outside of the Tokyo urban area. No significant differences were found in responses between the control and the experimental group.

### **Peer feedback task**

The peer feedback task took place during a practical phonetics course that met once a week for fifteen weeks in a CALL classroom equipped with iBuffalo-BSHSH12WH headsets. A typical 90-minute weekly lesson had the following sequencing: Timed fluency reading, short lecture on a concept in phonetics concept, individual listening and pronunciation practice of the targeted feature, peer feedback task, immediate post-test, and focused fluency practice. Each week, the lesson focused on different segmental or intonation contrasts; the present study uses data from the lessons on singleton-/l/•/ɪ/ and /l/•/ɪ/-clusters.

The peer feedback task and immediate post-test for the contrasts considered in the present study took place during the seventh (singleton-/l/•/ɪ/,  $n = 77$ ) and eighth (/l/•/ɪ/-clusters,  $n = 75$ ) class meeting. Table 1 shows the sequencing of the tasks and tests, and the data extracted for the present experiment.

**Table 1. Sequencing of Tasks and Tests.**

Week	Component	Task	Data used
2	Pre-test	Pronunciation	lamp, ramp, glass, grass
		Listening discrimination	light, right, lamp, ramp, lock, rock, glass, grass, play, pray, cloud, crowd
7	Individual training	Listening and pronunciation practice of singleton-/l/•/ɹ/	
	Whole-class training	Contrastive conversation for singleton-/l/•/ɹ/	
	Treatment	FtoF or CMC peer feedback on singleton-/l/•/ɹ/	
	Post-test	Pronunciation of singleton-/l/•/ɹ/	lamp, ramp
8	Individual training	Listening and pronunciation practice of /l/•/ɹ/-clusters	
	Whole-class training	Contrastive conversation for /l/•/ɹ/-clusters	
	Treatment	FtoF or CMC peer feedback on /l/•/ɹ/-clusters	
	Post-test	Pronunciation of /l/•/ɹ/-clusters	glass, grass
14	Delayed post-test	Pronunciation	lamp, ramp, glass, grass
		Listening discrimination	light, right, lamp, ramp, lock, rock, glass, grass, play, pray, cloud, crowd

The peer intelligibility feedback task used 8 minimal pairs for singleton-/l/•/ɹ/ (*leader/reader, lake/rake, lamb/ram, lamp/ramp, lice/rice, light/right, long/wrong, lock/rock*) and 8 for /l/•/ɹ/-clusters (*play/pray, blush/brush, cloud/crowd, clown/crown, glass/grass, glamour/grammar, flame/frame, flute/fruit*). Except for *rake, lice, blush, glamour, and clown*, the words were within the first 5000 words of the JACET8000, a wordlist commonly used in Japan for confirming word difficulty and the majority were at the Japanese junior high school level (Stewart & Stewart, 2016). These words appeared in minimal sentence conversations (e.g., Gilbert, 1993; Grant, 2007; Hancock, 2003) with L1 glosses. The first lines contained a phoneme contrast that elicited a different response, as seen in the following example:

Conversation 1 A: He is a good reader. B: Everyone trusts him.

Conversation 2 A: He is a good reader. B: He loves books.

Although not communicative, such drills require listeners to process meaning (Isaacs, 2009) and provide speakers with immediate evidence of conversational breakdown due to production or listening comprehension (Murphy & Baker, 2015).

Before beginning the peer feedback task, the participants were familiarized with the contrastive conversations through choral repetition. Participants in the FtoF group then formed self-selected pairs, having been instructed to avoid choosing the same partners each week. Participants took

turns reading the first line of one of the paired conversations in the textbook. The speaker's partner judged which of the two first lines had been uttered and responded with the next line. Participants used task sheets to record whether their partner's response was correct or not, then changed partners and repeated the process until the time limit was reached or until the task sheets were completed.

The CMC group used software developed for the purpose of delivering feedback from multiple online peers. The software randomly displayed the first line of one of the eight contrasting conversations. Each participant recorded this sentence in MP3 format, a format that has been found to be acceptable for intelligibility training using personal computers (Adachi, Akahane-Yamada, & Yamada, 2008). The software collected all recordings in a pool, and these were randomly delivered to other CMC participants who selected from a drop-down window which of the two responses they believe the utterance required. In order to deliver adequate feedback for each recording, the software was designed so that participants responded to three peers' recordings before receiving the next sentence to record. At the end of the feedback session, the software displayed to each participant a list of the sentences that they had recorded, along with the number of peers that had (and had not) chosen the correct response sentence, thus providing peer feedback on intelligibility. After confirming that all participants were on task, the instructor/researcher remained behind the instructor's console to avoid influencing the treatment outcome.

### **Immediate post-test**

After all participants had completed the feedback task, they took the online immediate pronunciation post-test. The computer screen displayed one set of contrastive conversations with L1 glosses and instructed participants to record the first line. Thus, for the present study, the participants produced four sentences (Week 7 *The lamp/ramp is broken*; Week 8 *Don't walk on the glass/grass*). Written Japanese instructions told participants to record only once and to keep recording even if they made a mistake. There were no questions.

### **Pre- and delayed post-tests**

The pre-test and delayed post-test for listening discrimination consisted of 24 pairs of words for eight phoneme contrasts taught in class. Listening discrimination data for the present study comes from six of these pairs: *light/right*, *lamp/ramp*, *lock/rock*, *glass/grass*, *play/pray*, *cloud/crowd*. Each word was embedded in the carrier phrase "I said [target word]". The target sentences were recorded with a Macintosh internal microphone in a slightly baffled environment. The talker, a female U.S. West Coast native-speaker with experience as a professional announcer and as an EFL instructor, used careful speech. The items were put in a different random order for the pre- and delayed post-test.

The pre- and delayed post-test for pronunciation consisted of 16 single-word prompts testing the 8 phoneme contrasts taught during the course: *base/vase*, *sink/think*, *lamp/ramp*, *glass/grass*, *cat/cut*, *hall/hole*, *boss/bus*, *heat/hit*. (For the present study, only the recordings of *lamp/ramp* and *glass/grass* were used.) Eight randomized prompt lists were created, four for the pre-test and four for the delayed post-test, while ensuring that members of the same minimal pair did not occur in succession.

Pre- and delayed post-tests were administered during the 2nd and 14th class meeting. During the pre-test, there were two questions about technical matters, and due to technical difficulties one participant took the listening test after the pronunciation test. Due to absences, three participants took the pre-test a week later, and one participant took the delayed post-test a few days later.



The listening test was played over the classroom loudspeaker to prevent participants from replaying the sound file on their computers. Participants listened to recorded instructions in English while reading them on the screen, and then were presented with a sample problem and were allowed to ask questions if necessary. Participants saw the two possible words onscreen and selected the word they had heard. For the pronunciation test, two versions of the prompt lists were distributed in alternating order so that participants did not sit next to another participant who had the same version of the test. Directions were read aloud by the researcher at the beginning of the test and were also displayed on the computer screen. Participants were instructed not to practice the words and to record each word twice at natural speed. This repetition allowed participants to self-correct and adjust their voice level if necessary. The tests took approximately 25 minutes to administer and there were no questions.

### **Rating of utterances**

The pre- and delayed post-pronunciation test data were prepared by extracting each participant's second utterance of *lamp*, *ramp*, *glass*, or *grass*. If the second utterance was unusable due to noise, the first utterance was used. The immediate post-test sound files were prepared by extracting the target word *lamp*, *ramp*, *glass*, or *grass* from the recorded sentence. For each target word, the audio files were randomly placed into sets. Each audio file was embedded into an online item with written multiple-choice responses of "L, R, Other" or "GL, GR, Other".

Two male native-speaker (Canada, U.S.) EFL instructors served as paid raters. A protocol was set up to ensure that the sets were rated in varying orders and that the raters listened to the two contrasts (singleton or cluster) in alternation. Raters did not know whether a given utterance was from the pre-, post-, or delayed post-test, or what the intended target phoneme was. The rating sessions took place in the CALL room where the course had taken place; raters were able to adjust the computer volume for each utterance and to replay utterances as necessary. Because the focus was on local intelligibility rather than global word or sentence intelligibility, raters were instructed to focus on the word-initial phoneme(s). They were told to choose the sound they had heard, selecting "other" only when this was not possible. For example, if the rater heard the target word *ramp* as /ɪʌm/, although the word is unintelligible, the initial /ɪ/ is locally intelligible, and so the rater would select R as the response. Four 60 to 90-minute sessions were held on four days. The first session began with oral and written instructions and a short familiarization practice. The raters worked at their own pace, alternating 20-minute rating sessions with 5-minute mandatory breaks until they completed the work.

Ratings were treated as binomial categories (rater heard target phoneme = intelligible; rater heard contrasting phoneme or "other" = not intelligible). The 956 audio files rated by both raters had 80% inter-rater reliability but a weak (.56) Cohen's kappa. Because the raters were well-trained and not likely to guess because they had the option of marking "other", 80% reliability is believed to indicate adequate reliability (McHugh, 2012). When the raters did not agree whether an utterance was intelligible, the researcher served as a third rater following the same protocol as the raters.

### **Analysis**

After the initial analysis, Generalized Estimating Equations (GEE) were run separately for each of the pronunciation targets (*lamp*, *ramp*, *glass*, *grass*). GEE produces a population-average model that describes the covariance among repeated observations for each subject without attempting to explain the source of this covariance. GEE was chosen because it can accommodate repeated-measures data with categorical outcomes (Heck, Thomas, & Tabata, 2012), and because it does not discard cases with missing values. The outcome variable was intelligibility or

unintelligibility of the target utterance. The within-subjects factor was time, measured as an ordinal variable; the between-subjects factor was group (FtoF or CMC).

Three grand-mean centered (Heck, Thomas, & Tabata, 2012) covariates were included in the models: Initial ability to discriminate /l/ and /ɫ/, as indicated by the score ranging from 0 to 12 for the 12 items on the listening pre-test (“Listening”); change in listening discrimination ability, as indicated by the difference in score between the listening pre-test and delayed post-test (“Listening Change”); and initial pronunciation confidence as indicated by the mean responses on a 5-point Likert scale to the eight can-do pronunciation statements on the demographic survey (“CanDo”).

When specifying the GEE models, robust standard errors were used as they are considered to be consistent even if the correlational structure is specified incorrectly (Heck, Thomas, & Tabata, 2012). Because the outcome variable had binary responses, a binomial distribution with a logit-link function was specified (Ballinger, 2004; Heck, Thomas, & Tabata, 2012). For each of the targets, two series of ten models were built from bottom-up by adding covariates and interactions. Little difference was found between exchangeable (CS) correlation structures and auto-regressive AR(1) structures. Models with CS structures and the lowest QICC were chosen. An alpha level of .05 was used for all statistical tests.

## Results

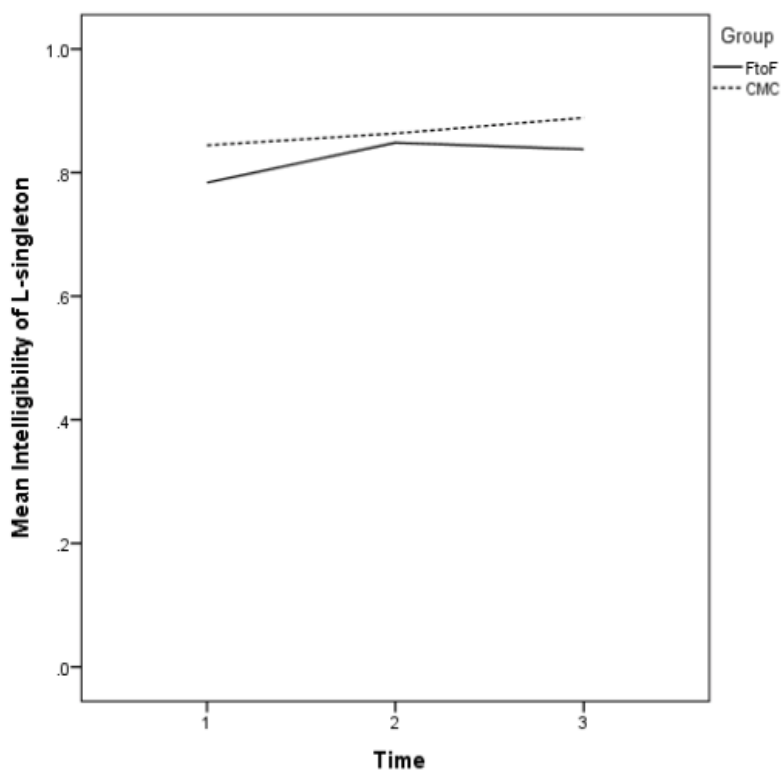
### Singleton-/l/

For both groups, singleton-/l/ was highly intelligible at the pre-test (Table 2), possibly introducing a ceiling effect. Mean intelligibility was significantly higher for the CMC group than for the FtoF group at the pre-test,  $t(74) = 4.01, p < .001, d = .93$  and at the delayed post-test,  $t(74) = 5.82, p < .001, d = 1.35$ , indicating that the CMC group was more intelligible before the treatment and remained so. The trajectories suggested steady improvement over time for the CMC group, but not for the FtoF group (Figure 1).

**Table 2. Mean Intelligibility of Singleton-/l/ by Group and Time.**

Group	Time	<i>M</i>	<i>SD</i>	Skewness	<i>SD</i>	Kurtosis	<i>SD</i>
FtoF <i>n</i> = 32	Pre	.78	.07	-1.43	.41	.04	.81
	Immed Post	.84	.07	-1.99	.41	2.08	.81
	Delayed Post	.81	.07	-1.68	.41	.88	.81
CMC <i>n</i> = 44	Pre	.84	.06	-1.93	.36	1.81	.70
	Immed Post	.86	.05	-2.20	.36	2.95	.70
	Delayed Post	.89	.05	-2.52	.36	4.56	.70

\*Based on cases with no missing values



**Figure 1. Mean Intelligibility of Singleton-/l/ by Group and Time.**

A GEE analysis of the singleton-/l/ data ( $n = 241$  from 83 participants) was conducted to determine the relationship between group and intelligibility. The most parsimonious model (QICC = 213.01) included Time and Group as factors. A significant main effect was found only for Intercept (Wald  $\chi^2(1) = 57.41, p < .01$ ) (Table 3).

**Table 3. Tests of Model Effects for Singleton-/l/.**

(Source)	Wald $\chi^2$	df	Type III Sig.
Intercept	57.407	1	.000
Group	.492	1	.483
Time	1.477	2	.478

Table 4 provides the parameter estimates. The odds ratio (*OR*) of the intercept indicated that participants' utterances were almost four times ( $\text{Exp}(\beta) = 3.80, p < .01$ ) more likely to be intelligible than unintelligible at the beginning of the study. The non-significant  $\text{Exp}(\beta)$  (odds ratio) for Time indicates that there was no significant change in the odds of being proficient at Time 2 or Time 3, and the non-significant  $\text{Exp}(\beta)$  (odds ratio) for Group indicates that there was no significant difference in the odds of being proficient for either group. There was no significant impact from any of the covariates.

**Table 4. Parameter Estimates for Singleton-/ɪ/.**

Parameter	$\beta$	SE	Hypothesis test			Exp( $\beta$ )	95% Wald CI for Exp( $\beta$ )	
			Wald $\chi^2$	df	Sig.		Lower	Upper
(Intercept)	1.34	.35	14.19	1	.00	3.80	1.90	7.62
[CMCgroup]	.32	.45	.49	1	.48	1.37	.57	3.33
[FtoFgroup]	0 <sup>a</sup>					1		
[Time=3]	.37	.32	1.34	1	.25	1.45	.77	2.71
[Time=2]	.31	.40	.61	1	.44	1.37	.62	2.99
[Time=1]	0 <sup>a</sup>					1		
(Scale)	1							

a. Set to zero because this parameter is redundant.

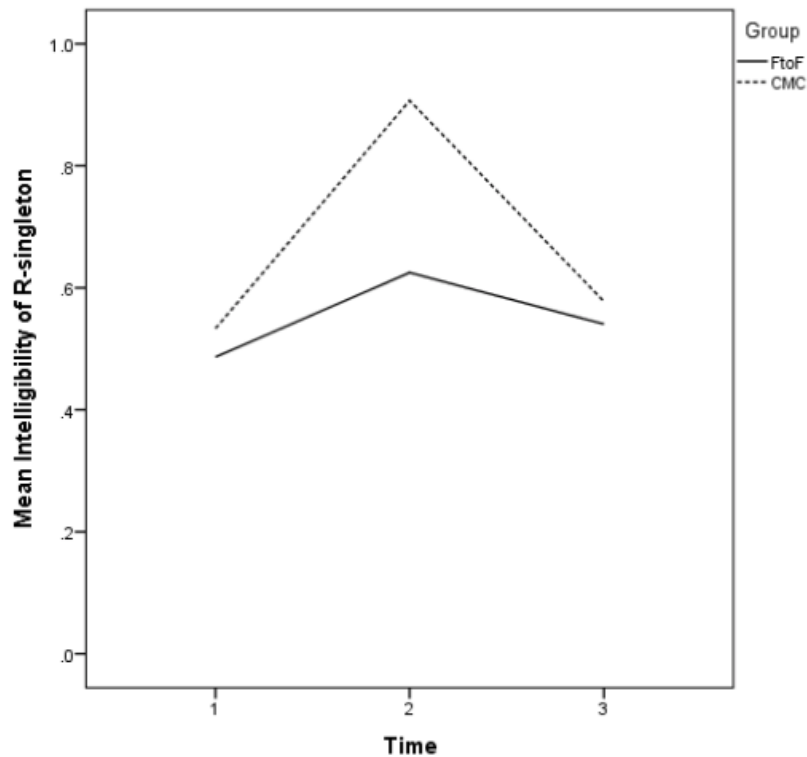
### Singleton-/ɪ/

The ability of participants in both groups to pronounce singleton-/ɪ/ intelligibly was much lower than for singleton-/l/, with about half of participants' being intelligible at the pre-test, and only slight improvement on the delayed post-test (Table 5). There was a significant difference in the means of the two groups at the pretest,  $t(72) = 4.53, p < .0001, d = 1.07$ ; at the immediate post-test,  $t(72) = 15.87, p < .001, d = 3.74$ , and at the delayed post-test,  $t(72) = 5.03, p < .001, d = 1.19$ , with large or very large effect sizes. Figure 2 visually confirms the large group difference at the immediate post-test.

**Table 5. Mean Intelligibility of Singleton-/ɪ/ by Group and Time.**

Group	Time	<i>M</i>	<i>SD</i>	Skewness	<i>SD</i>	Kurtosis	<i>SD</i>
FtoF <i>n</i> = 31	Pre	.42	.09	.34	.42	-2.02	.82
	Immed Post	.65	.09	-.64	.42	-1.71	.82
	Delayed Post	.48	.09	.07	.42	-2.14	.82
CMC <i>n</i> = 43	Pre	.51	.08	-.05	.36	-2.10	.71
	Immed Post	.91	.05	-2.91	.36	6.75	.71
	Delayed Post	.58	.08	-.34	.36	-1.98	.71

\*Based on cases with no missing values



**Figure 2. Mean Intelligibility of Singleton-/ɹ/ by Group and Time.**

A GEE analysis of the singleton-/ɹ/ data ( $n = 239$  from 83 participants) was conducted to determine the relationship between group and intelligibility. The most parsimonious model (QICC = 279.29) included Time and Group as factors and CanDo and Listening as covariates. Significant main effects were found for Intercept, Time, CanDo, and Listening (Table 6).

**Table 6. Tests of Model Effects for Singleton-/ɹ/.**

Source	Wald $\chi^2$	Type III df	Sig.
(Intercept)	9.473	1	.002
Group	1.236	1	.266
Time	19.237	2	.000
CanDo	4.628	1	.031
Listening	13.066	1	.000

Table 7 provides the parameter estimates. The OR for Time 2 indicated that, overall, participants were more than five times ( $\text{Exp}(\beta) = 5.03$ ,  $p < .01$ ) more likely to be intelligible at Time 2 than at the beginning of the study. However, the non-significant OR at Time 3 indicates that these gains did not carry over to the delayed post-test. The significant OR for CanDo indicates that for every one-unit increase in initial pronunciation confidence, almost two units of increased intelligibility could be predicted ( $\text{Exp}(\beta) = 1.95$ ,  $p = .03$ ). The significant OR for Listening indicates that for every one-unit increase in listening discrimination ability at the beginning of the study, over 1.5 units of increased intelligibility could be predicted ( $\text{Exp}(\beta) = 1.51$ ,  $p < .01$ ). In other words, for

singleton-/ɪ/, participants' intelligibility was impacted by their overall confidence in pronunciation as well as by their ability to discriminate it from /l/.

**Table 7. Parameter Estimates for Singleton-/ɪ/.**

Parameter	$\beta$	SE	Hypothesis test			Exp( $\beta$ )	95% Wald CI for Exp( $\beta$ )	
			Wald $\chi^2$	df	Sig.		Lower	Upper
(Intercept)	-.21	.34	.40	1	.53	.81	.42	1.56
[CMCgroup]	.44	.40	1.24	1	.27	1.56	.71	3.40
[FtoFgroup]	0 <sup>a</sup>					1		
[Time=3]	.23	.24	.99	1	.32	1.26	.80	2.00
[Time=2]	1.62	.37	18.61	1	.00	5.03	2.42	10.48
[Time=1]	0 <sup>a</sup>					1		
[CanDo]	.67	.31	4.63	1	.03	1.95	1.06	3.57
[Listening]	.41	.11	13.87	1	.00	1.51	1.22	1.88
(Scale)	1							

a. Set to zero because this parameter is redundant.

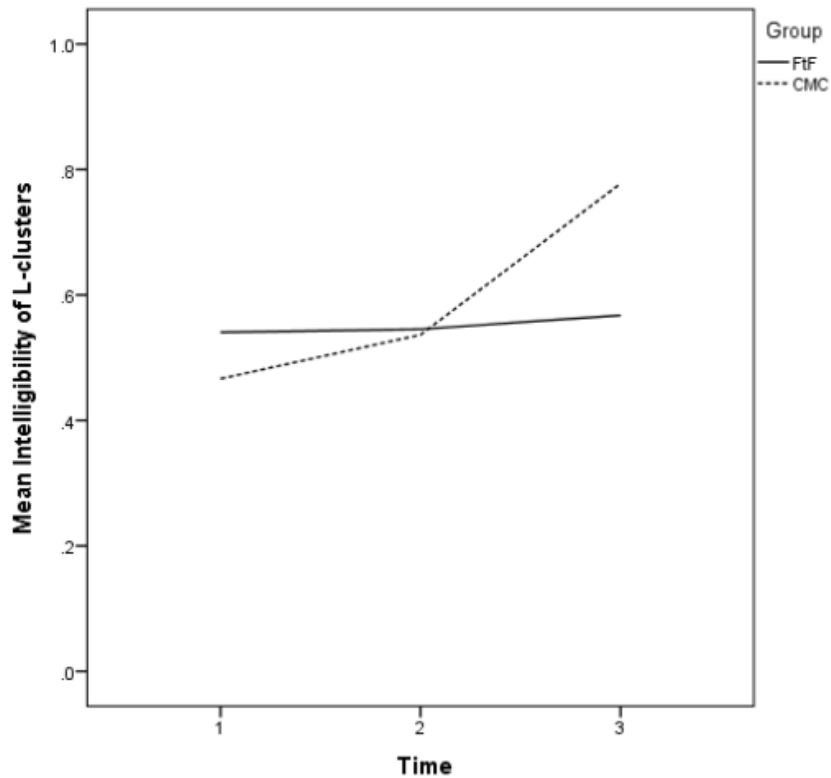
#### /l/-clusters

About half of the participants in both groups produced /l/-clusters intelligibly at the pre-test (Table 8). There was a significant difference in the mean scores of the two groups at the pretest,  $t(72) = 3.03, p = .0034, d = 0.71$ ). There was no statistical difference in means on the immediate posttest. The mean scores of the two groups were significantly different at the delayed posttest,  $t(72) = 8.06, p < .0001, d = 1.90$ ). The trajectories in Figure 3 show that the F-to-F group had little change throughout the treatment, but the CMC group made strong gains on the delayed post-test.

**Table 8. Mean Intelligibility of /l/-clusters by Group and Time.**

Group	Time	M	SD	Skewness	SD	Kurtosis	SD
FtoF <i>n</i> = 33	Pre	.55	.09	-.19	.41	-2.09	.80
	Immed Post	.55	.09	-.19	.41	-2.09	.80
	Delayed Post	.61	.09	-.46	.41	-1.91	.80
CMC <i>n</i> = 41	Pre	.49	.08	.05	.37	-2.01	.72
	Immed Post	.54	.08	-.15	.37	-2.08	.72
	Delayed Post	.76	.07	-1.24	.37	-.49	.72

\*Based on cases with no missing values



**Figure 3. Mean Intelligibility of /l/-clusters by Group and Time.**

A GEE analysis of the /l/-clusters data ( $n = 238$  from 83 participants) was conducted to determine the relationship between group and intelligibility. The most parsimonious model (QICC = 323.12) included Time, Group, and Time\*Group interaction as factors and Listening as a covariate. Significant main effects were found for Time and Time\*Group (Table 9).

**Table 9. Tests of Model Effects for /l/-clusters.**

Source	Wald $\chi^2$	Type III df	Sig.
(Intercept)	2.821	1	.093
Group	.389	1	.533
Time	9.540	2	.008
Listening	2.346	1	.126
Time*Group	6.787	2	.034

Table 10 shows the parameter estimates for /l/-clusters. The odds of the CMC group being intelligible at Time 3 were nearly four times as large as the odds of the FtoF group ( $\text{Exp}(\beta) = 3.69, p = .01$ ).

**Table 10. Parameter Estimates for /l/-clusters.**

Parameter	$\beta$	SE	Hypothesis test			Exp( $\beta$ )	95% Wald CI for Exp( $\beta$ )	
			Wald $\chi^2$	df	Sig.		Lower	Upper
(Intercept)	.17	.33	.27	1	.61	1.19	.62	2.28
[CMCgroup]	-.31	.45	.48	1	.49	.73	.30	1.77
[FtoFgroup]	0 <sup>a</sup>					1		
[Time=3]	.11	.33	.11	1	.74	1.12	.58	2.15
[Time=2]	-.04	.25	.03	1	.87	.96	.58	1.57
[Time=1]	0 <sup>a</sup>					1		
[Listening]	.15	.10	2.35	1	.13	1.16	.96	1.41
[Time3*CMCgroup]	1.30	.50	6.76	1	.01	3.69	1.38	9.85
[Time3*FtoFgroup]	0 <sup>a</sup>					1		
[Time2*CMCgroup]	.32	.45	.52	1	.47	1.38	.58	3.32
[Time2*FtoFgroup]	0 <sup>a</sup>							
[Time1*CMCgroup]	0 <sup>a</sup>							
[Time1*FtoFgroup]	0 <sup>a</sup>							
(Scale)	1							

a. Set to zero because this parameter is redundant.

**/ɹ/-clusters**

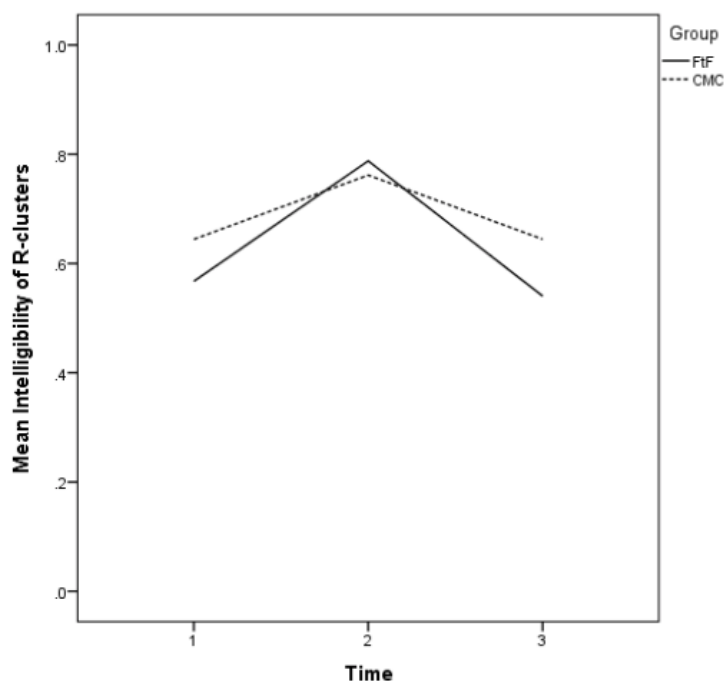
About 60% of participants pronounced /ɹ/-clusters intelligibly at the pre-test and delayed post-test (Table 11), a slightly higher rate than /ɹ/-singleton. The groups' mean scores were significantly different on the pretest,  $t(73) = 3.56, p = .0034, d = 0.48$ ) and on the delayed posttest,  $t(73) = 2.03, p = .0456, d = .48$ ), but not on the immediate post-test. The trajectories for both groups showed some improvement at the immediate post-test (Figure 4), returning to the pre-test levels at the delayed post-test.

**Table 11. Mean Intelligibility of /ɹ/-clusters by Group and Time.**

Group	Time	M	SD	Skewness	SD	Kurtosis	SD
FtoF <i>n</i> = 33	Pre	.55	.09	-.19	.41	-2.09	.80
	Immed Post	.79	.07	-1.48	.41	.19	.80
	Delayed Post	.58	.09	-.32	.41	-2.02	.80
CMC <i>n</i> = 42	Pre	.62	.08	-.51	.37	-1.83	.72
	Immed Post	.76	.07	-1.28	.37	-.39	.72
	Delayed Post	.62	.08	-.51	.37	-1.83	.72

\*Based on cases with no missing values





**Figure 4. Mean Intelligibility of /ɪ/-clusters by Group and Time.**

A GEE analysis of the /ɪ/-clusters data ( $n = 239$  from 82 participants) was conducted to determine the relationship between group and intelligibility. The most parsimonious model (QICC = 282.87) included Time and Group as factors and CanDo and Listening as covariates. Table 12 provides the tests of model effects. There were significant main effects for Intercept (Wald  $\chi^2(1) = 16.25, p < .01$ ), Time (Wald  $\chi^2(2) = 7.99, p = .02$ ) and Listening (Wald  $\chi^2(1) = 18.17, p < .01$ ).

**Table 12. Tests of Model Effects for /ɪ/-clusters.**

	Type III		
Source	Wald $\chi^2$	df	Sig.
(Intercept)	16.250	1	.000
Group	.968	1	.325
Time	7.994	2	.018
CanDo	1.491	1	.222
Listening	18.168	1	.000

Table 13 provides the parameter estimates for /ɪ/-clusters. The odds of being intelligible at Time 2 were more than twice as high ( $\text{Exp}(\beta) = 2.45, p = .01$ ) than at Time 1. However, the odds at Time 3 were slightly, but not significantly, lower than at Time 1, indicating that the gains made at the immediate post-test disappeared at the delayed post-test. The odds ratio for Listening indicated that for every one-unit increase in listening discrimination ability at the beginning of the study, over 1.4 units of increased intelligibility could be predicted ( $\text{Exp}(\beta) = 1.43, p < .01$ ), meaning that participants' ability to discriminate /ɪ/-clusters affected their intelligibility.

**Table 13. Parameter Estimates for /ɹ/-clusters.**

Parameter	$\beta$	SE	Hypothesis test			Exp( $\beta$ )	95% Wald CI for Exp( $\beta$ )	
			Wald $\chi^2$	df	Sig.		Lower	Upper
(Intercept)	.30	.31	.92	1	.34	1.35	.73	2.50
[CMCgroup]	.37	.38	.97	1	.33	1.45	.69	3.04
[FtoFgroup]	0 <sup>a</sup>					1		
[Time=3]	-.06	.25	.05	1	.82	.94	.57	1.55
[Time=2]	.90	.35	6.64	1	.01	2.45	1.24	4.84
[Time=1]	0 <sup>a</sup>					1		
[CanDo]	.38	.31	1.49	1	.22	1.46	.80	2.69
[Listening]	.36	.08	18.17	1	.00	1.43	1.22	1.69
(Scale)	1							

a. Set to zero because this parameter is redundant.

## Discussion

The first research question asks to what extent modality of feedback (FtoF or CMC) affects intelligibility. The results for /l/-clusters showed a significant effect for the CMC group at the delayed post-test, indicating that CMC feedback was more effective. The CMC group showed significant improvement, but the FtoF group showed none, a striking result given both groups' relatively high level of intelligibility for singleton-/l/. A possible explanation is that, as listeners and feedback-givers, the FtoF group did not have the time to notice and process the auditory qualities of /l/-clusters. As speakers, they were unable to transfer their ability to produce intelligible singleton-/l/ to a cluster environment. On the other hand, it seems that the CMC group was able to build on their reasonably high level of intelligibility for singleton-/l/ because the CMC environment provided them with adequate time to focus on acoustic cues. They could not immediately produce /l/-clusters more intelligibly at the immediate post-test, but they absorbed sufficient information to begin noticing the acoustic features of /l/-clusters, enabling them to improve their intelligibility after the feedback session. This is supported by findings from education in general that the cognitive benefits of peer feedback can accrue as "sleeper effects" after the assessment takes place (Topping, 1998). The results for /l/-clusters suggest that CMC peer feedback on intelligibility may be more useful than FtoF feedback if learners have a threshold level of competence for that target.

As for the other three targets, the results are less clear. For these three targets, the CMC group had a statistically higher group mean at the pre-test and delayed post-test: the CMC group began with an advantage over the FtoF group and maintained that advantage at the delayed post-test. Any immediate gains in intelligibility for participants in the FtoF mode disappeared by the delayed post-treatment, indicating that, overall, the FtoF treatment was less effective than the CMC treatment.

During peer feedback sessions, the FtoF participants displayed satisfaction at being understood by their peers, for example, by thanking peers when they provided the intended response. Likewise,

they expressed surprise or disappointment when they realized that their peers had not received the intended message. Despite these affective benefits, the results of this study do not indicate that FtoF feedback is effective for fostering intelligibility. It can be inferred that, for peer feedback on intelligibility (as opposed to feedback on comprehensibility negotiated by different-L1 peers during tasks), what is important is not so much the receiving of feedback, which is bound to be inflated due to the ISIB, but the giving of feedback, especially when learners have time to notice acoustic features. This study suggests that learners may receive more immediate benefits when feedback is delivered in a CMC modality; however, these benefits of CMC mode are moderated by target variability and to a lesser extent, by learner factors.

The second research question asks to what extent target variability affects intelligibility. We have already discussed /l/-clusters. Singleton-/l/ represents a target that is generally intelligible to NS raters and, due to ISIB, is also highly intelligible for participants. Learners' intelligibility continues to improve with time regardless of environment, but improvement is not significant, supporting the findings of Aoyama et al. (2004). In sum, singleton-/l/ is relatively stable, and same-L1 peer feedback in either the FtoF or CMC environment is insufficient to nudge learners into noticing features that would increase intelligibility. This suggests that same-L1 peer feedback on intelligibility, in either mode, is ineffective for targets that are already mostly intelligible. For such targets, peer negotiated feedback on comprehensibility during task-based activities might be more effective.

The results for both singleton-/ɪ/ and /i/-clusters confirm that Japanese speakers require long-term, intensive perceptual and pronunciation training for these targets. In the present study, the treatment was too short for either group to make significant delayed post-test gains on either target. However, participants as a whole did make strong gains at the immediate post-test, with parameter estimates showing singleton-/ɪ/ to be five times more likely, and /i/-clusters more than twice as likely, to be intelligible than at the beginning of the study. Additionally, although the GEE did not show a time\*group interaction, the group mean for singleton-/ɪ/ was significantly higher for the CMC group than for the FtoF group. Although native speakers attend to variance in the third formant (F3) to perceive the /l-/ɪ/ contrast, Japanese learners ignore this variance and pay attention to F2 frequencies (Saito, 2013). It is likely that the CMC group benefitted from having time to focus on this acoustic information.

These results indicate that, to some extent, short-term same-L1 peer feedback on intelligibility can lead to immediate uptake when delivered by CMC. As receivers of feedback, learners realize that their utterances are not necessarily intelligible, even to their peers who have an advantage of knowing the same L1, and as givers of feedback, they must make a conscious distinction between well-formed and intermediate forms. However, further research is needed to confirm whether same-L1 peer feedback that is given during a longer treatment would be sufficient for long-term uptake for a wider variety of targets. Same-L1 peer feedback may trigger learners' awareness, but it is not likely to be the most efficient way to create a new perceptual category.

The third research question asks to what extent learner variables affect intelligibility. An interesting finding of this study is that the covariates affected the results for /ɪ/, but not /l/. First, for singleton-/ɪ/ and /i/-clusters, the effect of peer feedback was moderated by listening discrimination ability at the beginning of the study. This implies that, for difficult phonemes such as /ɪ/, a threshold level of competence in listening discrimination is necessary for same-L1 peer feedback to be effective. Further research in this area could provide practical guidance for classroom teachers wishing to implement same-L1 peer feedback.

For singleton-/ɪ/, the effect of peer feedback was also moderated by initial pronunciation confidence, indicating that socio-psychological factors are at play as well. The Japanese accent is positioned negatively, and Japanese learners are highly concerned about the intelligibility of their pronunciation (Chiba, Matsuura & Yamamoto, 1995; Galloway, 2011; Matsuda, 2003; Matsuura, Chiba & Hilderbrandt, 2001). Japanese learners consider the ability to pronounce /ɪ/ correctly to be the hallmark of good pronunciation. In short, for singleton-/ɪ/, peer feedback might also be informed by the social status associated with an English /ɪ/ and the stigma associated with Japanese /ɾ/.

The participants provided peer feedback on a variety of phonemes and intonation patterns throughout the semester, but only two treatment sessions focused on the targets in this study. Thus, these results should be interpreted with some caution. The decision to limit the targeted items to four words was made for practical reasons, given the relatively large number of participants. Another weakness is that the four target words had varying familiarity as measured by JACET (*lamp*-Level 2, *ramp*-Level 5, *glass*-Level 2, *grass*-Level 1), which may have affected the outcome (Flege, Takagi & Mann, 1996), despite familiarizing the participants with the target items during in-class practice. Finally, the effect of feedback might have been stronger if this had been a laboratory study, as laboratory-based pronunciation instruction studies have been found to produce stronger effects than intact-class classroom studies (Lee, Jang, & Plonsky, 2015).

Pronunciation research tends to focus on NS-NNS interaction or on interaction among different-L1 learners, and relatively little is known about the efficacy of same-L1 peer pronunciation feedback. The present paper adds some evidence that, in a CMC environment, same-L1 peer feedback on intelligibility can promote pronunciation development. Further research is needed to confirm how same-L1 learners can best learn pronunciation from and through each other, much as they do in other areas of language learning.

## About the Author

**Suzanne M. Yonesaka** is a professor at Hokkai-Gakuen University in Sapporo, Japan.

## References

- Adachi, T., Akahane-Yamada, R., & Yamada, T. (2008). Influences of speech compression on perception and learning of English speech by native speakers of Japanese. *Educational Technology Research, 31*(1-2), 41–48. <https://doi.org/10.15077/etr.KJ00005101194>
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /ɾ/ and English /l/ and /r/. *Journal of Phonetics, 32*(2), 233–250. [https://doi.org/10.1016/S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)
- Bahanshal, D. A. (2013). The effect of large classes on English teaching and learning in Saudi secondary schools. *English Language Teaching, 6*(11), 49–59. Retrieved from <http://www.ccsenet.org/journal/index.php/elt/article/view/31107>
- Baker, A. A. (2011). *Pronunciation pedagogy: Second language teacher cognition and practice* (Doctoral dissertation). Georgia State University, Atlanta, Georgia. Retrieved from [http://scholarworks.gsu.edu/alesl\\_diss/16](http://scholarworks.gsu.edu/alesl_diss/16)

- Baker, A. & Burri, M. (2016). Feedback on second language pronunciation: A case study of EAP teachers' beliefs and practices. *Australian Journal of Teacher Education*, 41(6), 1–19. Retrieved from <http://ro.uow.edu.au/sspapers/2449/>
- Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, 7(2), 127-150. Retrieved from <https://journals.sagepub.com/doi/10.1177/1094428104263672>
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114(3), 1600-1610. <https://doi.org/10.1121/1.1603234>
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, 61(5), 977–985. Retrieved from <https://link.springer.com/article/10.3758%2FBF03206911>
- Bueno-Alastuey, M. C. (2011). Synchronous-voice computer-mediated communication: Effects on pronunciation. *CALICO Journal*, 28(1), 1–20. Retrieved from <https://journals.equinoxpub.com/index.php/CALICO/article/view/23003>
- Bueno-Alastuey, M. C. (2013). Interactional feedback in synchronous voice-based computer mediated communication: Effect of dyad. *System*, 41(3), 543-559. <https://doi.org/10.1016/j.system.2013.05.005>
- Chang, C.-H. (2012). *Instruction on pronunciation learning strategies: Research findings and current pedagogical approaches* (Master's thesis). University of Texas at Austin, Austin, Texas. <http://hdl.handle.net/2152/19930>
- Chiba, R., Matsuura, H., & Yamamoto, A. (1995). Japanese attitudes toward English accents. *World Englishes*, 14(1), 77-86. <https://doi.org/10.1111/j.1467-971X.1995.tb00341.x>
- Correa, M., & Grim, F. (2014). Audio recordings as a self-awareness tool for improving second language pronunciation in the phonetics and phonology classroom: Sample activities. *Currents in Teaching & Learning*, 6(2), 55–63. Retrieved from <https://www.worcester.edu/Currents-Archives/>
- Counselman, D. (2010). *Improving pronunciation instruction in the second language classroom* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3436059)
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16. <https://doi.org/10.1017/S0272263197001010>
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–398. Retrieved from <https://www.jstor.org/stable/3588486>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.
- Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, 30(2), 155-166. [https://doi.org/10.1016/S0346-251X\(02\)00012-X](https://doi.org/10.1016/S0346-251X(02)00012-X)
- Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, 41(1), 25–37. <https://doi.org/10.1016/j.system.2013.01.005>

- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 1(1), 3–18. Retrieved from <http://repositories.cdlib.org/uccllt/12/vol1/iss1/art2/>
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–424. Retrieved from <https://www.jstor.org/stable/3588487>
- Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɪ/ and /I/. *The Journal of the Acoustical Society of America*, 99(2), 1161-1173. <https://doi.org/10.1121/1.414884>
- Foote, J. A., Holtby, A. K., & Derwing, T. M. (2012). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, 29(1), 1–22. <https://doi.org/10.18806/tesl.v29i1.1086>
- Franklin, A. D., & Stoel-Gammon, C. (2014). Using multiple measures to document change in English vowels produced by Japanese, Korean, and Spanish speakers: The case for goodness and intelligibility. *American Journal of Speech-Language Pathology*, 23(4), 625–640. Retrieved from <https://ajslp.pubs.asha.org/article.aspx?articleid=1886811>
- Galloway, N. (2011). *An investigation of Japanese university students' attitudes towards English* (Doctoral dissertation). Faculty of Humanities, University of Southampton. Retrieved from <https://eprints.soton.ac.uk/345128/>
- Gilakjani, A., Ahmadi, S., & Ahmadi, M. (2011). Why is pronunciation so difficult to learn? *English Language Teaching*, 4(3), 74–83. Retrieved from <http://www.ccsenet.org/journal/index.php/elt/article/view/11877>
- Gilbert, J. (1993). *Clear speech*. Cambridge: Cambridge.
- Grant, L. (2007). *Well said (Introduction)*. Boston: Thomson.
- Hancock, M. (2003). *English pronunciation in use*. Cambridge: Cambridge.
- Hattori, K. (2010). *Perception and production of English /r/-/l/ by adult Japanese speakers* (Doctoral dissertation). University College London. Retrieved from <http://discovery.ucl.ac.uk/19204/>
- Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, 119, 1740. <https://doi.org/10.1121/1.2166611>
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel modeling of categorical outcomes using IBM SPSS*. NY: Routledge.
- IIBC (Institute for International Business Communication). (June 2017). *TOEIC program data and analysis 2017: 2016 nendo juken shasuu to heikin sukoo [testee numbers and mean scores in 2016 academic year]* [Pamphlet]. Tokyo: IIBC.
- Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review*, 64(4), 555–580. <http://dx.doi.org/10.3138/cmlr.64.4.555>
- Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, 27(1), 1–12. <https://doi.org/10.18806/tesl.v27i1.1034>

- Ishikawa, T. (2015). *A study of Japanese university students' attitudes towards their English* (Doctoral dissertation). School of Modern Languages, University of Southampton. Retrieved from <https://eprints.soton.ac.uk/394667/>
- Kashiwagi, A., & Snyder, M. (2008). American and Japanese listener assessment of Japanese EFL speech: Pronunciation features affecting intelligibility. *The Journal of AsiaTEFL*, 5(4), 27–47.
- Kim, S. M. & Lee, C. H. (2015). A case study on online peer feedback for learning English pronunciation at university level. *Multimedia-Assisted Language Learning*, 18(4), 42–69. Retrieved from [http://journal.kamall.or.kr/wp-content/uploads/2016/01/Kim+Lee\\_18\\_4\\_02.pdf](http://journal.kamall.or.kr/wp-content/uploads/2016/01/Kim+Lee_18_4_02.pdf)
- Knoch, U. (2017). What can pronunciation researchers learn from research into second language writing? In T. Isaacs & P. Trofimovich, P. (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 54–66). Bristol: Multilingual Matters.
- Lansing, C. R. & McConkie, G. W. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, and Hearing Research*, 42(3), 526–539. Retrieved from <https://jslhr.pubs.asha.org/article.aspx?articleid=1780974>
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Lefkowitz, N. & Hedgcock, J. (2002). Sound barriers: Influences of social prestige, peer pressure and teacher (dis)approval on FL oral performance. *Language Teaching Research*, 6(3), 223–244. <https://doi.org/10.1191/1362168802lr107oa>
- Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94, 1242. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3509365/>
- Loewen, S., & Isbell, D. R. (2017). Pronunciation in face-to-face and audio-only synchronous computer-mediated learner interactions. *Studies in Second Language Acquisition*, 6, 1–32. <https://doi.org/10.1017/S0272263116000449>
- Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals*, 41(2), 364–379. <https://doi.org/10.1111/j.1944-9720.2008.tb03297.x>
- Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning*, 29(3), 451–476. <https://doi.org/10.1080/09588221.2014.963123>
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1–40. <https://doi.org/10.1017/S0261444812000365>
- Matsuda, A. (2003). The ownership of English in Japanese secondary schools. *World Englishes*, 22(4), 483–496. <https://doi.org/10.1111/j.1467-971X.2003.00314.x>
- Matsuura, H. (2007). Intelligibility and individual learner differences in the EIL context. *System*, 35(3), 293–304. <https://doi.org/10.1016/j.system.2007.03.003>
- Matsuura, H., Chiba, R., & Hilderbrandt, P. (2001). Beliefs about learning and teaching communicative English in Japan. *JALT Journal*, 23(1), 67–82. Retrieved from [http://jalt-publications.org/recentpdf/jj/2001a\\_JJ.pdf#page=66](http://jalt-publications.org/recentpdf/jj/2001a_JJ.pdf#page=66)

- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25(3), 481–520. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.2307/3586981>
- Munro, M. J., & Derwing, T. M. (2015a). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, 1(1), 11–42. <https://doi.org/10.1075/jslp.1.1.01mun>
- Munro, M. J., & Derwing, T. M. (2015b). Intelligibility in research and practice: Teaching priorities. In M. Reed and J. M. Levis (Eds.), *The handbook of English pronunciation*, (pp. 377–396). Malden, MA: Wiley Blackwell. <https://doi.org/10.1002/9781118346952.ch21>
- Murphy, J. M. & Baker, A. A. (2015). History of ESL pronunciation teaching. In M. Reed and J. M. Levis (Eds.), *The handbook of English pronunciation*, (pp. 36–65). Malden, MA: Wiley Blackwell. <https://doi.org/10.1002/9781118346952.ch3>
- O’Neal, G. (2013). No need to quit your flapping: The intelligibility of flap /r/ phoneme substitutions for either the /ɹ/ or /l/ phonemes in non-native English speaker conversations. *Niigata Studies in Foreign Languages and Cultures*, 18, 39–62.
- O’Neal, G. (2015). Interactional intelligibility: The relationship between consonant modification and pronunciation intelligibility in English as a Lingua Franca in Japan. *Asian Englishes*, 17(3), 222–239. <https://doi.org/10.1080/13488678.2015.1041871>
- Otsuka, T. & Ueda, H. (2011). Chuugaku koukou de no hatsuon gakushuu rireki to teichaku dou [Pronunciation instruction and its achievements in Japanese junior and senior high schools]. *Osaka Jogakuin Daigaku Kiyou*, 8, 1–27.
- Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly*, 34(4), 711–737. <https://doi.org/10.2307/3587782>
- Roccamo, A. (2015). Teaching pronunciation in just ten minutes a day: A method for pronunciation instruction in first-semester German language classrooms. *Die Unterrichtspraxis/Teaching German*, 48(1), 59–83. <https://doi.org/10.1111/tger.10181>
- Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, 35(1), 1-29. <https://doi.org/10.1017/S0272263112000666>
- Saito, K. (2014). Experienced teachers’ perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics*, 24(2), 250–277. <https://doi.org/10.1111/ijal.12026>
- Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3, 243–261. <https://doi.org/10.1017/S0142716400001417>
- Sicola, L. (2009). *No, they won’t ‘just sound like each other’: NNS-NNS negotiated interaction and attention to phonological form on targeted L2 pronunciation tasks*. Frankfurt am Main: Peter Lang.
- Stewart, J., & Stewart, W. (2016). Using the JACET 8000 word list to evaluate L2 vocabulary and reading difficulty. *Journal of Chikushi Jogakuen University*, 11, 57–66.



- Sueyoshi, A. & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661–699. <https://doi.org/10.1111/j.0023-8333.2005.00320.x>
- Swerts, M. & Kraemer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, 36(2), 219–238. <https://doi.org/10.1016/j.wocn.2007.05.001>
- Szypra-Kozłowska, J. (2014). *Pronunciation in EFL instruction*. Bristol, U.K.: Multilingual Matters.
- Timson, S., Grow, A., & Matsuoka, M. (1999). Error correction preferences of second language learners: A Japanese perspective. *JACET Bulletin*, 30, 135–148.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. <https://doi.org/10.3102/00346543068003249>
- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(01), 122–140. <https://doi.org/10.1017/S1366728914000832>
- Walker, R. (2001). Pronunciation priorities, the Lingua Franca Core, and monolingual groups. *Speak Out*, 18, 4–9. Retrieved from <https://associates.iatefl.org/pages/materials/ltskills19.doc>
- Walker, R. (2005). Using student-produced recordings with monolingual groups to provide effective, individualized pronunciation practice. *TESOL Quarterly*, 39(3), 550–558. <https://doi.org/10.2307/3588495>
- Walker, R. & Zoghbor, W (2015). The pronunciation of English as a lingua franca. In M. Reed and J. M. Levis (Eds.), *The handbook of English pronunciation*, (pp. 433–453). Malden, MA: Wiley Blackwell. <https://doi.org/10.1002/9781118346952.ch24>
- Wang, H., & van Heuven, V. J. (2015). The Interlanguage Speech Intelligibility Benefit as bias toward native-language phonology. *i-Perception*, 6(6), 1–13. <https://doi.org/10.1177/2041669515613661>
- Yoon, S. Y., & Lee, C. H. (2009). A study on voice recordings and feedback through BBS in teaching and learning pronunciation. *Multimedia-Assisted Language Learning*, 12(2), 187–216. Retrieved from <http://kamall.or.kr/kor/publications/MALL/12-2-2009.pdf#page=187>
- Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36(1), 69–84. <https://doi.org/10.1016/j.system.2007.11.004>