

Learner perceptions of online peer pronunciation feedback through P-Check

Suzanne M. Yonesaka

Hokkai-Gakuen University
syonesaka@gmail.com

Receiving adequate pronunciation feedback is an ongoing challenge for L2 learners. Although instructors are the most important source of corrective pronunciation feedback (Szpyra, 2014; Timson, 2007), L2 learners can also benefit from peer pronunciation feedback (Lord, 2008; Kim & Yoon, 2014; Roccamo, 2015). This paper examines Japanese university students' perceptions of online peer feedback on intelligibility (Derwing & Munro, 2005). To aggregate the feedback, P-Check, a plug-in for the Glexa LMS (Version2, 2015) was developed and its recognition and dialog functions were piloted in two studies. In the first exploratory study (N = 59), a post-treatment survey and diaries were used to explore participants' attitudes. Participants believed that peer feedback through P-Check was useful and had increased their awareness of their pronunciation strengths and weaknesses. In the second quasi-experimental study (N = 55), no significant difference was found in participants' perceptions of the usefulness of peer feedback, whether delivered face-to-face or via P-Check. Recommendations for further development of P-Check and other similar plug-ins are given.

Keywords: pronunciation, peer feedback, intelligibility, CALL

Introduction

An ongoing challenge for language learners is the difficulty in receiving adequate feedback during focused pronunciation practice, especially in large pronunciation courses where instructors may be unable to provide adequate individualized corrective feedback (Baker, 2011). By accessing additional sources of feedback, learners can gain

a more accurate picture of how their pronunciation is perceived and understood by others. The present study, which focuses on feedback given by peers during focused phonemic practice sessions, examines learners' perceptions of pronunciation feedback by multiple peers that is delivered online.

In a meta-analysis of 86 quantitative experimental or quasi-experimental studies on the effect of pronunciation instruction, Lee, Jang, and Plonsky (2014) found that explicit pronunciation instruction can be effective for learners at all proficiencies. Their meta-analysis also confirmed the results of Saito's (2012) synthesis of 15 experimental classroom studies which found that pronunciation instruction is effective for improving both phonemic and supra-segmental aspects of pronunciation. Thus, focused phonemic practice that is appropriately delivered and that provides effective feedback should be one component of pronunciation instruction.

Two main sources of pronunciation feedback for the learner are the teacher and, more recently, the computer. Providing corrective feedback has been commonly considered to be one of the pronunciation teacher's most important responsibilities (Szypra-Kozłowska, 2014), and corrective feedback that is provided immediately and on a one-to-one basis by the teacher is a significantly more powerful teaching tool than listening-only activities (Dlaska & Krekeler, 2013). Another source of effective corrective feedback is computer-assisted pronunciation training and assessment (CAPTA) (Eskenazi, 2009; Levis, 2007; Neri, Cucchiaroni, Strik, & Boves, 2002; Tsurutani, 2008; Tsurutani et al., 2006), although it has also been argued that existing CAPTA software is often not as successful as it could be due to inaccurate automatic speech recognition (Tsurutani, 2008) or a lack of pedagogy in the feedback (Engwall, 2006).

Learners' peers can provide another valuable source of pronunciation feedback (Derwing, 2008; Morley, 1991). When reviewing 82 stand-alone pronunciation classes in Canada, Foote, Holtby, and Derwing (2010) found nearly 40 ways of assessing pronunciation, with peer evaluation being one of the most popular types. The term itself was not clearly defined, and likely included corrective feedback, global assessment, and focused evaluation of specific pronunciation features.

One reason for incorporating peer feedback in pronunciation pedagogy is that it allows both students to focus on the form. Lyster, Saito, and Sato (2013) found that in L2 speaking classrooms, not only students who received but also those who provided peer corrective feedback on pronunciation during spoken classroom interaction had an increased focus on form in pronunciation.

The field of pronunciation teaching is now moving toward empowering the learner toward autonomy through the development of "speech awareness, self-observation skills, and self-monitoring skills with a positive attitude" (Chang, 2012, p. 9). Another reason to use peer feedback is its potential to increase learners' ability to more accurately assess their own pronunciation. By recording their speech, learners have ample opportunity for critical listening to their peers' and their own speech and to notice features that they otherwise would not (Gilakjani & Ahmadi, 2011). Trofimovich, Isaacs, Kennedy, Saito, and Crowther (2016) note that because pronunciation feedback is often ambiguous or absent, learners are often unable to self-assess their pronunciation skills accurately. In other fields, peer-assessment has been recommended as one way of improving self-assessment skills, but its effectiveness in the area of pronunciation has only begun to be tested.

Compared to other skills, there is a paucity of classroom-based research on pronunciation in general (Baker & Murphy, 2011), and "peer assessment of pronunciation is an area

that is underexplored in the research literature and certainly an area in need of research” (Knoch, 2017, p. 66). Some studies describe the incorporation of pronunciation feedback in traditional classrooms without online elements. Walker (2005) used peer feedback for learners sharing the same L1, focusing on targeted features of English as an International Language. In an experimental study to examine the effectiveness of pronunciation instruction with 25 beginning students of German, Roccamo (2015) incorporated peer corrective and evaluative feedback both inside and outside the class. Although peer feedback was not the focus of either of these studies, the authors indicated that it was a critical component of instruction.

At least one study has compared the effect of different types of feedback on pronunciation. Ahangari (2014) investigated the effect of teacher, self, and peer correction on the pronunciation improvement of 45 Iranian EFL learners in oral production tasks. For fifteen sessions, the learners told a story based on a series of pictures, and their performance was recorded and the pronunciation was scored. Participants in the self-correction group received written feedback from their teacher indicating incorrectly pronounced words, but were required to find and practice the correct pronunciation of the words themselves. Participants in the peer correction group gave corrective feedback to each other. Participants in the teacher correction group received corrective oral feedback from the teacher during the task. Results found that the self-correction group received the highest post-test scores, the peer correction group received the second highest, and the teacher correction group received the lowest. However, it can be argued that because the teacher indicated pronunciation problems to the self-correction group, this group was less autonomous than the peer correction group. The study does indicate that peer correction, even among EFL learners with the same L1, is a viable option for autonomy in the pronunciation classroom.

The literature, although sparse, suggests pedagogical reasons why it may be worthwhile to use peer feedback in a CMC environment. One recent study found positive linguistic effects of peer feedback. Kim and Lee (2015) investigated the patterns of corrective peer feedback of ten Korean university non-English majors in an online pronunciation class, using observations of the classroom BBS, voice recordings, reflective learning journals, and interviews. Students produced 14 weekly voice recordings by repeating the sample recordings 30 times and then generating and uploading the final recording. Students had two days to listen to their group members’ recordings and give peer feedback according to a rubric. The researchers found that the learners produced more indirect peer feedback on supra-segmental features than segmental features, but that there was more uptake for the segmentals. The researchers concluded that giving peer feedback as well as receiving feedback contributed to improved production because the learners had to consciously perceive and identify L2 pronunciation features.

Other studies have examined learners’ attitudes regarding peer feedback in a CMC environment. Kim and Yoon (2014) used a questionnaire and interview to analyze the attitudes of 71 non-English majors taking an online pronunciation class at a Korean university toward corrective peer pronunciation feedback. Although the nature of how the feedback was given is not clear from their report, the learners reported peer corrective feedback to be effective and useful. Another study (Lord, 2008) involved collaborative podcasting with peer feedback as part of a Spanish phonetics class. The researcher had 16 university Spanish learners work in small groups to record six podcasts on particular oral tasks, with other group members providing online written pronunciation feedback directly on the podcast site. Although providing written feedback was one of the participants’ least favorite aspects

of the project due to the time involved, they valued receiving peer feedback and felt that it was useful and constructive.

Research in other skill areas suggests that conducting peer feedback online, as opposed to face-to-face, could also provide affective benefits to learners. Pronunciation is deeply linked to learner identity (Gatboton, Trofimovich, & Magid, 2005), so it is intrinsically an area in which learners feel anxious about giving or receiving feedback. Some learners find it socially unacceptable to give or receive peer feedback on spoken discourse (Lyster, Saito, & Sato, 2013) and on pronunciation (Suarez & Tanaka, 2001). Feedback delivered through an asynchronous CMC environment could go some way to alleviating this anxiety, as in other skills areas. In the teaching of writing, for example, it has been found that an asynchronous CMC environment can promote collaboration among students while reducing their inhibitions towards communication (Caws, 2006). Ho and Savignon (2007) examined the attitudes of 37 English majors in Taiwan regarding peer review of essays or short research papers, conducted face-to-face for the first semester and computer-mediated for the second. Most learners reported feeling more comfortable giving online feedback because they found it somewhat stressful to review others' work face to face. In the same way, it can be posited that asynchronous computer-mediated pronunciation feedback will be less threatening to students' public image than face-to-face feedback.

Having peers provide feedback online as opposed to face-to-face also increases pedagogical objectivity. In the same way that computer conferencing allows learners to concentrate on the message rather than the presenter because of the presenter's reduced social presence (Ocker & Yaverbaum, 1999), an asynchronous CMC environment allows the learner who is giving the feedback to focus on the utterance and not the speaker, who remains anonymous. The anonymity of the person giving the feedback is also critical because anonymous assessment encourages more truthful and appropriate attitudes toward the assessment processes (Wen & Tsai, 2006), resulting in fairer assessment (Freeman & McKenzie, 2002).

Despite these benefits, there are at least three possible drawbacks to peer pronunciation feedback when it is corrective. First, peer-produced corrective feedback in spoken interaction can be socially unacceptable, both to the receivers and the givers (Lyster, Saito, & Sato, 2013). Such an attitude was typical of Japanese students fifteen years ago, when a pronunciation attitude survey of 81 Japanese students found that only 25% had positive attitudes about practicing English pronunciation with peers (Suarez & Tanaka, 2001). The second drawback is that, in order to provide corrective feedback, learners need a certain level of knowledge of the linguistic feature (Lyster, Saito, & Sato, 2013); lacking this knowledge, learners can find it difficult to pinpoint problematic pronunciation areas (Kim & Yoon, 2014). A third drawback is that even learners who are aware of a pronunciation problem may not have enough phonetics background to explain what the problem is, and so provide incorrect or incomplete feedback (Kim & Lee, 2015). To mitigate these pitfalls, the present study focuses not on corrective peer feedback, but on peer feedback on intelligibility.

Intelligibility is "the extent to which a listener actually understands an utterance" (Derwing & Munro, 2005, p. 385). 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). The present study uses Derwing and Munro's (2005) definition of intelligibility, operationalizing it in a forced-choice judgment activity in which the hearer chooses which utterance was perceived.

In traditional pronunciation classes, feedback on intelligibility occurs during focused practice (e.g., Gilbert, 1984; Hancock, 2003) in which Student A utters one of two minimal pair sentences (e.g., *I need a red pencil. / I need a lead pencil.*) and Student B responds with the appropriate phrase (e.g., *Sorry, mine is black. / Sorry, mine is carbon.*) that demonstrates whether the original utterance was intelligible. The present study proposes that such focused practice will benefit by being conducted online. In computer-mediated communication, the lack of non-linguistic signals such as backchannels and facial expressions is often seen as a drawback, but in this case would be an advantage, allowing learners to focus exclusively on pronunciation. In addition, giving and receiving feedback with multiple peers in a CMC environment provides learners with more opportunities to focus on the targeted feature in a detached, laboratory-like environment, thus fostering the reflection and experimentation that are necessary for supporting learner autonomy (Schwienhorst, 2008).

In order to facilitate peer feedback on the intelligibility of specific segmental and supra-segmental features after explicit instruction by an instructor, the researcher collaborated with Version2 (2015) to develop P-Check, a learning management system (LMS) plug-in. The next section describes its functions.

Development of P-Check

P-Check is a plug-in for Glexa, a proprietary php-based LMS with a strong multi-media focus developed by Version2 that is presently used by 80 universities in Japan. P-Check's functions are: (a) random delivery of written prompts to learners, (b) recording, (c) random delivery of recordings to other learners for assessment, (d) collecting feedback on the quality of recording and on confidence in assessment, and (e) monitoring so that students and instructors see records of feedback given and feedback received.

P-Check presently supports three types of exercises: recognition, dialog, and dictation. For recognition exercises, the instructor creates a set of written prompts that differ in only one segmental or supra-segmental feature such as main sentence stress or intonation. P-Check randomly presents one of the written prompts onscreen for the learner to record, and then delivers the recording randomly to the learner's peers, the "P-Checkers". The original set of written prompts is also delivered to the P-Checkers in a drop-down menu. The P-Checkers listen to the recording and select the sentence they believe they heard, thus making a judgment regarding the intelligibility of the utterance. While doing so, the P-Checkers also indicate whether they are confident or not in their judgment.

The second type of exercise, dialog, requires students to choose the response to an utterance. To prepare the exercise, instructor writes two two-line dialogs in which the first line contains the contrasting feature that triggers a different response, as in this example:

1. A: It's a large base. B: There are many soldiers.
2. A: It's a large vase. B: We'll need a lot of flowers.

P-Check presents one of the first lines for the learner to record. The P-Checkers listen to the recording and select from a drop-down window the response they believe the utterance requires. Figure 1 shows a dialog exercise in which L1 glosses have been provided. The buttons for P-Checkers to indicate their confidence and to flag problematic recordings are also visible.

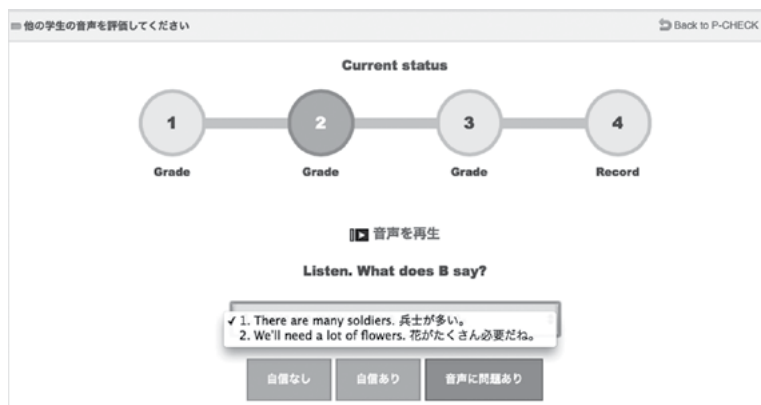


Figure 1. Interface when giving feedback on a dialog exercise.

Note: Translation of buttons: Not Confident, Confident, Problematic recording

For the third type of prompt, dictation, P-Check sends a written prompt (word, sentence, or longer utterance) to a learner who then reads and records it, after which it is randomly sent to the P-Checkers to transcribe. The dictation prompt function was not piloted in the present paper.

The first time learners access P-Check they receive a prompt to record, but subsequently they must give feedback to their peers by completing three P-Checks before receiving a prompt, ensuring adequate feedback from multiple P-Checkers. P-Checkers can flag problematic (e.g., low volume) recordings, which are eliminated from the system after three flags. After P-Checkers have recorded all of the items in an exercise, they can no longer record; however, they can continue to provide feedback. Accumulated feedback for each recording (number of times checked and % of P-Checkers perceiving the intended utterance) is displayed so that students can confirm the intelligibility of the targeted feature.

P-Check's design is kept deliberately simple, embodying the essential design features described by Boling and Soo (1999): consistency, good use of space, legibility, contrast, alignment, ease of navigation, high-quality playback. P-Check also meets most of the design elements for software in which learners have only moderate control (p. 453):

- ✧ The learning objectives are clear. (Yes.)
- ✧ The content is divided into relatively small units. (Yes.)
- ✧ The chunks are sequenced from simple to difficult. (No, prompts are delivered randomly, but the instructor can create P-Checks at varying levels of difficulty.)
- ✧ Feedback is provided constantly and instantly. (Yes.)
- ✧ Branching is available and is based on an adequate assessment tool. (No.)
- ✧ Navigation gives the learners the choice of what (No) and how fast (Yes) to learn.
- ✧ An anchor or bookmarking system brings learners back to the main pages. (Yes.)
- ✧ The learners are initiators of actions, including assessments. (Yes.)
- ✧ The software guides the learners through the content. (Yes.)

Although a focus-on-form perspective is usually associated with the acquisition of grammar forms, pronunciation can also be considered within this framework (Levy & Kennedy, 2004; Lyster, Saito, & Sato, 2013). P-Check requires learners to provide feedback indicating the presence or absence of an error, analogous to how conversational prompts indicate the presence of an error without providing an exemplar of the target form or an explanation of how the speaker should correct it. Conversational prompts aim to improve learners' control over forms by providing opportunities for pushed output (Lyster, Saito, & Sato, 2013). In other words, in conversation, a prompt would be normally followed by self-repair. With P-Check, the feedback should be followed by individual practice of the targeted feature before further output via P-Check. Because the recording and feedback are asynchronous, learners have adequate time to focus on form (Levy & Kennedy, 2004) and do sub-vocal rehearsal. P-Check provides multiple opportunities for learners to discriminate and also to produce the target structures, allowing them to restructure their knowledge of features of English pronunciation.

Research question

Although some research indicates that online peer feedback on pronunciation is effective, little research has been done in this area. However, research in other skill areas suggests that an asynchronous environment will reduce learners' anxiety by providing them with the time and anonymity to give objective pronunciation feedback. The present study focuses on learners' reactions to the giving and receiving of aggregated peer online feedback delivered through P-Check. After confirming the functionality of P-Check, the first step in assessing its pedagogical effectiveness usefulness is to gauge learner reactions (Levy, 2007). To that end, the primary research question guiding the present analysis is: To what extent do learners perceive online peer feedback to be effective in supporting their pronunciation development?

Method

This exploratory research paper presents partial results of two studies piloting P-Check in the spring and fall semesters of 2015. Study 1 pilots P-Check's recognition exercise and also explores participants' attitudes through a pre- and post-treatment survey. Study 2, which pilots P-Check's dialog exercise, is a quasi-experimental study comparing participants' reactions to peer feedback delivered face-to-face and via P-Check.

Participants

The two studies used two different groups of participants. The participants in Study One were 59 Japanese university students ($M = 35$, $F = 24$) of which 45 were first-year students, enrolled in the Spring 2015 semester of practical phonetics. The participants in Study Two were 65 Japanese university students ($M = 29$, $F = 36$) of which 48 were first-year students, enrolled in the Fall 2015 semester of practical phonetics.

All were Japanese university students majoring in Humanities at a large private mid-level university in Japan. All were native speakers of Japanese and none had spent more than 30 days in an English-speaking country. Their English language ability ranged from low intermediate to high intermediate. In a goal-setting activity at the beginning of the

course, the participants had been asked how often they had their pronunciation corrected by a teacher, with 30% responding never, 56% responding sometimes, and 14% responding often. This reflects the peripheral status of pronunciation teaching at Japanese junior and senior high schools (Oda, 2012; Tominaga, 2011), and along with anecdotal classroom evidence, indicates that as a whole, the participants had not received large amounts of pronunciation instruction. Informed consent was obtained in accordance with TESOL standards.

The participants were intact groups enrolled in an elective practical phonetics course, and as such, they are convenience samples. However, they typify the population of non-elite Japanese university participants as can be seen from this university's admissions results on the National Center Test for University Admissions, a standardized university entrance exam for public and private schools.

The practical phonetics course met weekly for 15 weeks for 90 minutes in a CALL room for Study 1 and in a computer laboratory equipped with lesser quality headsets for 15 weeks for Study 2. It used a bilingual practical phonetics textbook (Yonesaka & Tanaka, 2013), and three types of audio-visual instructor-produced materials were available on the Glexa LMS for classwork and independent study: (a) discrimination, production and P-Check exercises for targeted segmental and supra-segmental features, (b) International Phonetic Alphabet (IPA) reading flashcards, and (c) practices for weekly phonetics content and IPA recognition quizzes.

Study One

Instruments. Three instruments were used to explore participants' attitudes about pronunciation in general and about their perceptions of P-Check in particular: a pre-treatment survey, a post-treatment-survey, and learner pronunciation diaries. The pre-treatment survey used five items drawn from previous pronunciation attitude surveys (Derwing & Rossiter, 2002; Yoshida et al., 2012). Based on these surveys, five questions were constructed asking about pronunciation: whether learners had ever received feedback on their English pronunciation, how they would assess their own pronunciation of English sounds as compared to their peers, how they would assess their own fluency as compared to their peers, what their pronunciation goals were, and how deeply they understood the importance of pronunciation in communication. In response to Baker and Murphy's (2011) recommendation for more research on learners' perceptions of pronunciation feedback, two new items were created to elicit learners' reactions to feedback: *It is helpful to get feedback from my teacher about my pronunciation during class*, and *I feel comfortable getting feedback from other students about my pronunciation during class*. The survey also included five demographic questions about learners' gender, native language, length of time they had studied English before entering university, length of time they have spent in an English-speaking country, and how often they speak English for personal reasons.

The post-treatment survey included six of the seven attitude questions. Five additional questions about P-Check were created based on items from Lord's (2008) survey; however, the points on the Likert-type scale were increased from five to six to avoid neutral responses. Both surveys were translated into Japanese by a bilingual Japanese TESOL professional and checked by the researcher. (See Appendix for English versions.)

The learners were required to write weekly pronunciation diaries in English or in Japanese. Comments regarding P-Check were extracted and translated into English by the researcher for use in confirming survey results.

One instrument was used to assess learners' pronunciation. At the end of the course, learners practiced and recorded a 123-word passage written at the learners' reading level and constructed to elicit a variety of pronunciation features. The recording was rated by the instructor/researcher on ten segmental and supra-segmental features, and the composite score was used to assess their pronunciation.

Materials. The P-Check exercises that are the focus of Study One used three singleton consonant minimal pairs: /b/•/v/, /l/•/r/, and /s/•/θ/. (See Table 1.) These were chosen for their orthographic consistency and because these distinctions are difficult for Japanese learners, particularly /l/•/r/ (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999) which also carries a high functional load (Munro, 2011).

Table 1. Minimal pair prompts

Contrast	Prompts
/b/•/v/	(Abbie/Avie) made a (bow/vow). (Ebban/Evan) needs your (boat/vote).
/l/•/r/	(Lenny/Rennie) is a good (leader/reader). (Lina/Rina) is (collecting/correcting) the papers.
/s/•/θ/	(Selma/Thelma) has a big (mouse/mouth). (Ross/Roth) took the mountain (pass/path). (Bess/Beth) lost (face/faith) at the meeting.

The sentences in Table 1 were used to create sets of four written prompts, each containing two instances of the targeted phonemic contrast (e.g., *Abbie made a bow*, *Abbie made a vow*, *Avie made a bow*, and *Avie made a vow*). Although these prompts lack communicative value, they allow learners to notice the contrastive phonemes that affect communication. Prompts can be constructed for other features such as sentence stress and intonation for English language learners, pitch accent for Japanese language learners, and tones for Chinese language learners. As explained earlier, P-Check randomly presents one of the prompts on screen for the learner to record, and then delivers the recording randomly to the P-Checkers, along with the original set of written prompts in a drop-down menu. The P-Checkers listen to the recording and select the sentence they believe they heard, thus making a judgment regarding the intelligibility of the utterance.

Data collection. During the fourth and fifth class meeting, after learning about the characteristics of the targeted segmental features, the participants engaged in discrimination and production practice. During the discrimination exercises, participants heard twenty pairs of words, ten of which were different (e.g., *light*, *right*) and ten of which were tokens of the same word (e.g., *right₁*, *right₂*), marked these as same or different, and received a score. During the focused pronunciation exercises, participants heard and read onscreen, then shadowed, three sentences in which the targeted segmental features were made salient through colored fonts. The first sentence contained multiple instances of the first phoneme, the second contained the second phoneme, and the third contained both. Finally, participants heard and read onscreen, then shadowed and recorded, a one-paragraph story containing multiple instances of both phonemes similarly highlighted with colored fonts.

After completing the practices, participants used P-Check exercises for approximately 10 minutes during the fifth through ninth class meetings and could also access them for independent study. It should be noted that Glexa's recording function used Java Applet, and at the time of data collection, changes in Internet security triggered constant security warnings, a disruption in workflow that could not be resolved by university tech staff.

The pre-treatment survey was administered during the third and fourth lesson, and the post-treatment survey was administered during the last two lessons. Both surveys were administered via the Glexa LMS with no time constraints. A total of 51 participants completed both surveys and there were no questions or comments. A total of 59 paper-based pronunciation diaries were collected during the last class.

To prepare the 994 recordings that had been submitted for P-Checking, the researcher did a blind listening of each recording (i.e., researcher intelligibility judgment). The recordings were presented in random order and both the prompt and P-Check responses were hidden, so that the researcher did not know what the intended utterance was. At this point, the sound files could not be downloaded from the user interface, so a second rater could not be used for inter-rater reliability; instead, intra-rater reliability was determined. Six months later, the same researcher did a second intelligibility judgment of 168 (17%) of the recordings. Cohen's κ determined that there was strong agreement between the two researcher intelligibility judgments ($\kappa = .855, p = .000$).

In order to determine recording and error patterns, each recording was coded and annotated for recording quality, flagging, and confidence level using data available on the instructor feedback page. Of the 994 recordings, 85 (9.2%) had been flagged by at least one P-Checker, and of these problematic files, 40 had been uploaded by six participants. In other words, approximately half of the flagged files originated from the few participants who experienced ongoing technical difficulties.

To confirm the accuracy of the information displayed on the instructor page, the data were matched against the P-Check database and the coding was checked again, resulting in a total of 872 non-problematic recordings.

Study Two

Materials. Based on previous experience and on standard resources (Avery & Ehrlich, 2004; Dale & Poms, 1994; Kenworthy, 1992), eight phoneme contrasts that are typically problematic for Japanese speakers of English were the focus of this study: /b/•/v/, /s/•/θ/, initial /l/•/r/, initial /l/•/r/ clusters, /æ/•/ʌ/, /ɑ/•/ou/, /ɑ/•/ʌ/, and /i/•/ɪ/.

Materials for each minimal pair consisted of pairs of two-line contrastive conversations with L1 glosses. The first lines were the same except for a phoneme contrast that required a different response from Student B, as seen in the following example:

Conversation 1 Student A: He is a good leader. 彼は良い指導者だ。
Student B: Everyone trusts him. だれでも彼を信頼する

Conversation 2 Student A: He is a good reader. 彼は良い読者だ。
Student B: He loves books. 彼は本が大好き。

The contrastive conversations, which appeared in the textbook as well as in P-Check, were used in Steps 3 and 4 of the teaching module (see Treatment). Table 2 lists the minimal pairs that were the focus of the contrastive conversations. The /i/•/ɪ/ module was originally

taught in Lesson 8, but due to time constraints, could not be completed and was repeated in the Lesson 10. Data from Lesson 10 are used in this study.

Table 2. Minimal pairs featured in the contrastive conversations

Lesson	Feature	Minimal pairs
4	/b/•/v/	berry-very, base-vase, boat-vote, bow-vow, curb-curve
4	/s/•/θ/	sinking-thinking, mouse-mouth, face-faith, moss-moth, pass-path
5	/l/•/r/	long-wrong, light-right, lamp-ramp, lock-rock, lice-rice, leader-reader, lake-rake, lamb-ram
7	/æ/•/ʌ/	cap-cup, cat-cut, track-truck, bag-bug, ankle-uncle
7	/ɑ/•/oʊ/	ball-bowl, lawn-loan, saw-sew, walk-woke, hall-hole, pause-pose
8	/ɑ/•/ʌ/	caught-cut, shot-shut, lock-luck, jog-jug, boss-bus
9	/l/•/r/ clusters	playing-praying, flame-frame, cloud-crowd, flute-fruit, glass-grass, blush-brush, clown-crown, glamour-grammar
10	/i/•/ɪ/	feel-fill, heat-hit, leave-live, beat-bit, neat-knit

Treatment. The treatment took place from the fourth to tenth lesson (except Lesson 6) toward the end of each class (See Table 2). During Lesson 3 to 6, the instructor/researcher used the bilingual textbook and visual aids to teach about consonants (e.g. manner of articulation) and vowels (e.g. front, back). After both contrastive phonemes had been taught, the treatment began. For example, in Lesson 3 the participants learned about consonants vs. vowels, stops, and nasals. In Lesson 4 they learned about fricatives, at which point the contrastive /b-v/ treatment could begin. By Lesson 7, lecture content had moved on to syllables, but the participants continued with the treatment.

Before beginning the treatment each week, the instructor/researcher reviewed the characteristics of each phoneme, emphasizing how they contrasted. Immediately afterwards, the participants completed an online module consisting of five steps, of which Step 4 is the experimental treatment.

For each module, 10 to 16 different participants were placed in the experimental face-to-face feedback (F-to-F) group, and the remaining participants were assigned to the P-Check group. This ensured that all participants received and gave both types of feedback for consonants and vowels. To control for listening discrimination proficiency of the targeted phonemes in each module, the F-to-F participants were a stratified random sample based on the results of the listening discrimination test in Step 1.

Step 1. Listening discrimination pretest.

For this individual, online pretest, participants heard twenty minimal pairs, half of which were different words (e.g., *light* and *right*) and half of which were different tokens of the same word (e.g., *light*₁ and *light*₂), and marked each pair as the same or different. The materials used in Steps 1 and 2, which had been developed separately from the contrastive conversations, did not necessarily contain the same target contrasts as the conversations.

Step 2. Listening discrimination and pronunciation practice.

This individual, online practice contained four parts, in which color-coded typeface promoted noticing of the targeted contrasts: (a) shadowing ten minimal pairs, then shadowing

three sentences, the first with multiple occurrences of the first phoneme, the second with multiple occurrences of the second phoneme, and the third with multiple occurrences of both; (b) repeating the same materials as in Step 1 for listening practice; (c) watching a video of a native-speaker peer model pronouncing key words; (d) listening to, then recording, a short paragraph containing multiple examples of the phoneme contrast. During Step 2, the instructor/researcher assigned participants to the F-to-F group.

Step 3. Practice using the textbook exercise.

Participants did whole-class, choral practice of the contrastive conversations in the textbook to become familiar with them.

Step 4. Peer intelligibility feedback.

Participants used the same materials as in Step 3. Participants in the F-to-F group stood up, made pairs, and completed the textbook exercise face-to-face, rotating to new partners when they finished. The P-Check group used the same materials through P-Check.

Step 5. Follow-up.

Participants responded (Agree, Disagree, I don't know) individually online to three statements: (a) I can usually pronounce the two target sounds so that a listener will be able to tell the difference, (b) I can usually hear the difference between the two target sounds, and (c) The peer check (face-to-face or P-check) was useful.

The original 65 participants attempted between two and eight modules on the targeted features. Table 3 provides participant information for each module. Participants attempted an average of 1.5 modules using F-to-F peer feedback and an average of 5.3 modules using P-Check peer feedback.

Table 3. Number of participants attempting each module

Lesson	Targeted feature	F-to-F feedback	P-Check feedback	Total
4	/b/•/v/	15	38	53
4	/s/•/θ/	9	44	53
5	/l/•/r/	13	42	55
7	/æ/•/ʌ/	10	44	54
7	/ɑ/•/oʊ/	14	44	58
8	/ɑ/•/ʌ/	14	45	59
9	/l/•/r/ clusters	13	40	53
10	/i/•/ɪ/	10	47	57

Only data from participants who had completed five or more of the modules were included in this study. Of the original 65 participants, 55 completed five or more modules, for a total of 391 valid response sets. To be considered valid, the response set needed to include the listening score from Step 1, responses to at least two of the three statements in Step 5, and, for participants giving P-Check feedback, a verified P-Check recording.

40 Instruments. The 55 participants' attitudes regarding the usefulness peer feedback were measured through their responses to the third statement in Step Five at the end of each

module: *The peer check (face-to-face or P-Check) was useful.* The responses were collapsed into two (Yes; No or I don't know) in order to fill the chi-square cells during data analysis.

Results

Study one

The data in this section come from the post-treatment survey exploring learners' perceptions of P-Check. Participants responded to five statements about P-Check using a six-point Likert-type scale from 1 *strongly disagree* to 6 *strongly agree*. For each statement, a chi-square test of goodness-of-fit was performed to determine whether the six choices were equally preferred.

Two statements concern affective aspects: PC1 *I enjoyed the P-Check project* and PC4 *Recording my own voice makes me nervous.* (See Table 4.)

Table 4. Affective aspects of P-Check, by frequency and %

	Strongly disagree	Disagree	Disagree Slightly	Agree Slightly	Agree	Strongly agree	Total
Enjoyed	0 (0.0)	0 (0.0)	4 (7.8)	11 (21.6)	24 (47.1)	12 (23.5)	51 (100)
Nervous	3 (5.9)	5 (9.8)	4 (7.8)	14 (27.5)	18 (35.3)	7 (13.7)	51 (100)

Responses were not equally distributed in the population, either for PC1 *Enjoyed*, $X^2(3, N = 51) = 16.2, p = .001$, or for PC4 *Nervous*, $X^2(5, N = 51) = 21.8, p = .001$. Although most participants said that recording their own voice made them nervous, over 90% enjoyed P-Check to some extent, a finding that is corroborated by the weekly pronunciation diaries.

One statement concerned technical aspects: PC2 *Technological glitches with P-Check severely impeded the success of the project.* (See Table 5.)

Table 5. Technical Aspects of P-Check, by Frequency and %

	Strongly disagree	Disagree	Disagree slightly	Agree slightly	Agree	Strongly agree	Total
PC2							
Technological glitches	0 (0.0)	4 (7.8)	6 (11.8)	13 (25.5)	20 (39.2)	8 (15.7)	51 (100)

For PC2, responses were not equally distributed in the population $X^2(4, N = 51) = 16.2, p = .003$. About 80% of the participants felt to some extent that P-Check's success had been severely impeded by technological glitches; these were, in fact, internet security problems or equipment-related difficulties.

Two statements concern pronunciation awareness: PC3 *I became more conscious of my own pronunciation through this project (more so than I would have through just taking the phonetics course.)* and PC5 *I learned a lot about my own pronunciation strengths and weaknesses through P-Check.* (See Table 6.)

Table 6. Awareness-raising aspects of P-Check, by frequency and %

	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree	Total
PC3 Became conscious	0 (0.0)	0 (0.0)	0 (0.0)	8 (15.7)	20 (39.2)	22 (43.1)	50 (98)
PC5 Strengths & weaknesses	0 (0.0)	2 (3.9)	3 (5.9)	9 (17.6)	19 (37.3)	18 (35.3)	51 (100)

Responses were not equally distributed in the population for PC3, $X^2(2, N = 50) = 6.9, p = .032$, or for PC5, $X^2(4, N = 50) = 25.4, p = .000$. Over 80% of the respondents agreed or strongly agreed, and none disagreed, that P-Check had helped them become conscious of their own pronunciation. Over 90% of the participants agreed to some extent that P-Check had helped them learn about their own pronunciation strengths and weaknesses.

Relationships among P-Check survey questions. In further analysis, relationships among the individual P-Check survey questions were examined. A Spearman's correlation found a moderate statistical correlation between PC1 *Enjoyment* and PC2 *Glitches* ($r_s = .424; p = .002$). There was a fairly large effect size but the confidence interval (C.I.) was quite wide (95% C.I.: 0.169, 0.626; $N = 51; R_2 = .179$). PC1 *Enjoyment* was also found to correlate weakly with PC3 *Awareness*, ($r_s = .299; p < .05$). The effect size of the correlation was medium but the C.I. was quite wide (95% C.I.: 0.023, 0.532; $N = 50; R_2 = .089$). PC3 *Awareness* was found to correlate moderately with PC5 *Learned* ($r_s = .334, p < .05$). There was a medium effect size but the C.I. was quite wide (95% C.I.: 0.062, 0.56; $N = 50; R_2 = .111$).

Relationships among P-Check survey questions and learner variables. Next, the responses to the P-Check questions were examined in terms of learner variables. No relationship was found between participants' responses and their pronunciation goals at the end of the course.

Spearman's correlations found a moderate negative statistical correlation between post-course pronunciation scores and PC5 *I learned a lot about my own pronunciation strengths and weaknesses through P-Check*, ($r_s = .359, p = .010$), with a medium effect size (95% C.I.: -0.577, -0.093; $N = 51; R_2 = .128$). The more that participants felt they had learned about their pronunciation, the less accurate was their final pronunciation.

Relationships between P-Check survey questions and performance as P-Checker. Participants' responses to the P-Check survey questions were examined in terms of their performance as P-Checkers. The participants' role as P-Checker may be observed in terms of extent (how much they participated in giving P-Checks), effectiveness (how effectively they used the software), and accuracy (how accurately they were able to perform P-Checks).

First, because there was a wide range in the number of P-Checks that participants performed, from 11 to 79, it seems likely that their survey responses would be affected by the sheer amount of P-Checking they engaged in. A Spearman's correlation found a moderate statistical correlation between the number of P-Checks that participants did on their peers' recordings and PC2 *Technological glitches with P-Check severely impeded the success of the*

project, ($r_s = .298, p = .033$). The effect size of the correlation was medium but the C.I. was quite wide (95% C.I.: 0.025, 0.530; $N = 51, R_2 = .088$).

A moderate statistical correlation was also found between the number of P-Checks that participants did on their peers' recordings and PC5 *I learned a lot about my own pronunciation strengths and weaknesses through P-Check*, ($r_s = .320, p = .022$). However, the effect size of the correlation was small and the C.I. was wide (95% C.I.: 0.149, 0.547; $N = 51, R_2 = .102$), meaning that the correlation coefficient is not highly reliable.

Relationships between P-Check survey questions and performance as system user. To explore participants' perceptions of technological glitches, correlations (Spearman) examined the relationship between the frequency that participants' recordings were flagged and their responses to the P-Check survey questions. The frequency of flagged recordings correlated fairly strongly with PC2 *Technological glitches with P-Check severely impeded the success of the project* ($r_s = .52, p < .001$). The effect size of the correlation was very large and the C.I. was quite narrow, meaning that the correlation coefficient is highly reliable (95% C.I.: 0.288, 0.697; $r_s = .522, N = 51, R_2 = .272$). There is a strong and important relationship between the frequency that participants' recordings were flagged, and the severity with which they judged technological glitches.

Relationships between P-Check survey questions and performance as pronouncer. To investigate the relationship between participants' attitude toward P-Check and improvement in pronunciation accuracy, a new variable, Pronunciation Improvement, was calculated by subtracting pre-course pronunciation scores from post-course scores. A histogram of pronunciation improvement showed normal distribution and a boxplot revealed a symmetrical distribution except for three outliers. No significant correlations were found between participants' pronunciation improvement and their responses to the five items asking about P-Check. Likewise, no significant correlations were found between pronunciation improvement and the amount of P-Check use (number of prompts recorded, number of P-Checks received, and number of P-Checks performed) or the quality of use as measured by the number of flags received.

Although the participants had, overall, a very positive attitude toward receiving peer pronunciation feedback online, it is possible that similar results would have been found if the feedback had been given face-to-face, and an experimental study comparing the efficacy of the two types feedback is needed. As a preliminary step, a second quasi-experimental study was conducted to assess the effectiveness of both types of feedback while piloting P-Check's dialog function.

Study two

The data in this section consist of the participants' reactions to online or face-to-face peer feedback after completing each of the eight modules. Of the 391 responses to the statement *The peer check (face-to-face or P-Check) was useful*, 81.1% indicated agreement, with a slightly higher percentage of positive responses from the F-to-F participants. (See Table 7.)

Table 7. Responses to the statement that peer feedback is useful by frequency and %

Feedback type	Agree (%)	Disagree or don't know (%)	Total (%)
P-Check	246 (80.1)	61 (19.9)	307 (100)
Face-to-face	71 (84.5)	13 (15.5)	84 (100)
Total	317 (81.1)	74 (18.9)	391

Next, the relationship between the type of peer feedback and perceptions of the usefulness of peer feedback was assessed to confirm whether this difference was significant. Because each participant responded multiple times, a two-way group-independence chi-square was performed separately for each of the eight modules. No significant differences were found.

Discussion

This study investigated participants' reactions to online peer feedback on intelligibility. The main finding is that most participants overall felt that it was useful for their pronunciation development, and that it was as useful as feedback delivered face-to-face. The participants in the first study agreed that the feedback helped them become more conscious of their pronunciation, a finding supported by statements in the weekly pronunciation diaries, such as "I began P-Check, but I was shocked because I didn't readily have everybody hear my pronunciation." The feedback's narrow focus drew participants' attention to the impact of the unintelligibility of specific features on communication.

The participants in the first study also felt that online peer feedback helped them discover their own particular pronunciation strengths and weaknesses. Participants could be observed checking their feedback status, confirming the targeted feature, rehearsing it, and then returning to P-Check. Sometimes the discoveries occurred while providing feedback, as seen in this diary entry: "Listening to the practice of other Japanese has been useful for me to check my own pronunciation. In particular, my r sound is too exaggerated." A possible explanation for this participant's comment is that normally, during classroom communication tasks, EFL learners tend to converge toward pronunciation that is heavily influenced by their L1 (Levis, 2005). However, while practicing and recording targeted features, learners' attention is directed to the selected features, reducing the chance of converging on the L1 (Walker, 2005) and preserving a range of pronunciations. When giving P-Check feedback, the participants may have noticed a range of pronunciations by their classmates and compared them to their own.

The interaction between receiving peer feedback and grasping one's pronunciation strengths and weaknesses is not straightforward. One troubling finding of the first study is that the more that participants felt they had learned about their pronunciation strengths and weaknesses through peer feedback, the less accurate was their final pronunciation. One interpretation is that although the peer feedback encouraged poorer students, it was not sufficient for uptake, due to feedback inaccuracy or learner variables. Another finding was that participants' sense of having learned about their strengths and weaknesses correlated moderately with the number of P-Checks they performed, but the effect was small. Even a limited amount of peer feedback may be sufficient for learners to feel that they have learned about their strengths and weaknesses, regardless of whether such learning has occurred.

research is needed to confirm its actual efficacy. A discussion of the accuracy of the P-Checkers' intelligibility judgments is beyond the scope of this study, but judgments were not as accurate as had been hoped for. When participants chose from four choices in Study 1, the mean accuracy of their intelligibility judgments was 37%; when choosing from two choices in Study 2, the mean accuracy increased, but only to 64%. Clearly, more work is needed to determine the learner, prompt, and system variables that promote or impede the accuracy of feedback on intelligibility. In addition, research will need to examine learner uptake of the feedback, both immediate and long-term.

The second main finding was that P-Check was enjoyable for participants in the first study to use despite their ongoing nervousness while recording and despite occasional glitches. These participants maintained their initial positive attitudes regarding the importance of pronunciation and slightly changed their pronunciation goals, putting a stronger emphasis on intelligibility rather than achieving native speaker-like pronunciation. Because Study 1 is not a true experimental study, it is not possible to know whether the use of online peer feedback influenced participants' attitudes, but the shift in participants' pronunciation goals towards intelligibility is in line with P-Check's focus. Further experimental research is needed to determine if learners enjoy P-Check specifically, or whether they would find any form of online feedback equally enjoyable. Even so, it is encouraging that participants' attitudes toward P-Check were overwhelmingly positive, and they that believed P-Check increased their awareness of their own pronunciation strengths and weaknesses.

Limitations and suggestions for further development

There are several limitations of this exploratory study that need to be addressed. First and foremost, because this is an exploratory study, the findings are tentative and should be treated with caution. A true experimental study is needed to compare learners' attitudes toward, and uptake of, online and face-to-face peer feedback on intelligibility over time.

Second, the results are restricted to data from learners of only one L1 background (Japanese) with relatively little pronunciation training. It would be revealing to see whether advanced Japanese learners provide more accurate assessments. This study used minimal pair prompts that are a challenge for Japanese learners, but different prompts that target pronunciation features of other L1s may produce quite different results with other groups of learners. Future research in a variety of contexts will be useful to tease out effects from learner variables (e.g. L2 proficiency; discrimination proficiency) and effects from language variables (e.g. targeted features, wording of prompts).

Third, the attitudinal survey may not have been detailed enough to deeply probe learners' attitudes. Some of the items should be refined, and binary responses should be replaced with scales. Another needed refinement is in the P-Check interface that asks for a binary confidence judgment (confident/not confident). The interface was designed this way to help participants maintain a focus on the task at hand without feeling anxious or getting waylaid while determining their precise level of confidence. However, one of the reviewers suggested that, in the future, the implementation of a sliding scale would provide more accurate data both for student and research purposes.

Finally, this study piloted only two of P-Check's functions, so further investigation will be needed to confirm the smooth functioning of the dictation prompt, a function that will likely be useful for many teaching needs beyond pronunciation.

Despite the weaknesses of this study, the P-Check plug-in has strong potential for

promoting pronunciation and oral fluency. Rather than isolating students, it brings them together, if not quite to collaborate, at least to cooperate in pronunciation development, hopefully leading to increased learner autonomy and independence.

P-Check's development continued in 2016 based on results of this study. For example, an instructor page showing flagged files was added to enable instructors to assist students experiencing technical difficulties as early as possible. Sound files can now be downloaded so that multiple teachers or researchers will be able make intelligibility judgments. Meanwhile, the Glexa LMS audio has been changed from Java to Flash, greatly improving its performance, and instructors can now perform P-Checks directly into the system while students use it.

In the future, we hope to add a function that matches the instructor's intelligibility judgments against the P-Checkers' judgments, ranks individual P-Checkers in terms of judgment accuracy, and then weights the P-Checkers' judgments according to their rankings. If learners knew that their utterances and their intelligibility judgments were being monitored in real time, this would add an element of accountability to the system, making it more pedagogically sound. In addition, preliminary data show that there is quite a wide disparity in individual learners' ability to decode their peers' utterances, so the weighting of judgments should provide more accurate feedback both to the learner who recorded and to the P-Checkers. Because learners are accustomed to websites that feature ongoing rankings, the somewhat more game-like atmosphere would probably encourage learners to use P-Check for independent study as well as during class.

To conclude, pronunciation learners need multiple ways of giving and receiving pronunciation feedback, including feedback from their peers. With further refinements, P-Check should be a useful and unique tool to assist learners in their ongoing pronunciation development.

Acknowledgements

The development of P-Check was funded through the CALL committee at Hokkai-Gakuen University. I thank Akio Ohnishi of Version2 and my colleagues for their support: Yukie Ueno for her administrative and translation assistance and Hiroya Tanaka for his critical comments on an earlier draft of this paper.

References

- Ahangari, S. (2014). The effect of self, peer and teacher correction on the pronunciation improvement of Iranian EFL learners. *Advances in Language and Literary Studies*, 5(1), 81-88. <http://dx.doi.org/10.7575/aiac.all.v.5n.1p.81>
- Avery, P. & Ehrlich, S. (2004). *Teaching American English pronunciation*. Oxford, UK: Oxford University Press.
- Baker, A. (2011). *Pronunciation pedagogy: Second language teacher cognition and practice* (dissertation). Georgia State University, Atlanta, Georgia. http://scholarworks.gsu.edu/alesl_diss/16
- Baker, A. & Murphy, J. (2011). Knowledge base of pronunciation teaching: Staking out the territory. *TESL Canada Journal*, 28(2), 29-50. Retrieved from <http://www.teslcanadajournal.ca/index.php/tesl/article/view/1071/890>

- Boling, E. & Soo, K.-S. (1999). CALL issues: Designing CALL software. In J. Egbert & E. Hanson-Smith (Eds.), *CALL environments: Research, practice, and critical issues*. (pp. 442–456). Bloomington, IL: TESOL.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. <http://dx.doi.org/10.3758/BF03206911>
- Caws, C. (2006). Assessing group interactions on-line: Students' perspectives. *Journal of Learning Design*, 1(3), 19–28. <http://dx.doi.org/10.5204/jld.v1i3.23>
- 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>
- Dale, P., & Poms, L. (1994). *English pronunciation for Japanese Speakers*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Derwing, T. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen Edwards and M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347–369). Amsterdam: John Benjamins.
- Derwing, T. & Munro, M. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16. <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/>
- Derwing, T. & Munro, M. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <http://dx.doi.org/10.2307/3588486>
- Derwing, T. M. & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, 30(2), 155–156. [http://dx.doi.org/10.1016/S0346-251X\(02\)00012-X](http://dx.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. <http://dx.doi.org/10.1016/j.system.2013.01.005>
- Engwall, O. (2006). Feedback strategies of human and virtual tutors in pronunciation training. *Speech, Music and Hearing*, 48(1), 11–34. <http://www.speech.kth.se/qpsr>
- Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, 51(10), 832–844. <http://dx.doi.org/10.1016/j.specom.2009.04.005>
- Footte, J., Holthy, A., & Derwing, T. (2010). Survey of the teaching of pronunciation in adult ESL programs in Canada. *TESL Canada Journal*, 29(1), 1–22. <http://dx.doi.org/10.18806/tesl.v29i1.1086>
- Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects. *British Journal of Educational Technology*, 33(5), 551–569. <http://dx.doi.org/10.1111/1467-8535.00291>
- Gatboton, E., Trofimovich, P., & Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL Quarterly*, 39(3), 489–511. <http://dx.doi.org/10.2307/3588491>
- Gilakjani, A. & Ahmadi, M. (2011). Why is pronunciation so difficult to learn? *English Language Teaching*, 4(3), 74–83. <http://dx.doi.org/10.5539/elt.v4n3p74>
- Gilbert, J. (1984). *Clear speech*. Cambridge, UK: Cambridge University Press.
- Hancock, M. (2003). *English pronunciation in use*. Cambridge: Cambridge University Press.

- Ho, M.-C., & Savignon, S.J. (2007). Face-to-face and computer-mediated peer review in EFL writing. *CALICO Journal*, 24(2), 269–290. Retrieved from <http://www.jstor.org/stable/24147912>
- 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>
- Kenworthy, J. (1992). *Teaching English pronunciation*. London, UK: Longman.
- 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
- Kim, S. & Yoon, S. (2014). Students' perspectives on the use of peer feedback in the online English pronunciation class at university level [PowerPoint slides]. Paper presented at the Korea Association of Multimedia-Assisted Language Learning Annual Conference 2014. Seoul, Korea. Retrieved from <http://www.slideshare.net/engedukamall/kamall-2014-presentation-kim-yoon>
- 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. <http://dx.doi.org/10.21832/ISAACS6848>
- Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <http://dx.doi.org/10.1093/applin/amu040>
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <http://dx.doi.org/10.2307/3588485>
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184–202. <http://dx.doi.org/10.1017/S0267190508070098>
- Levy, M. (2007). Research and technological innovation in CALL. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 180–190. <http://dx.doi.org/10.2167/illt069.0>
- Levy, M., & Kennedy, C. (2004). A task-cycling pedagogy using stimulated reflection and audio-conferencing in foreign language learning. *Language Learning and Technology*, 8(2), 50–68. Retrieved from <http://llt.msu.edu/vol8num2/levy/>
- Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals*, 41(2), 364–379. <http://dx.doi.org/10.1111/j.1944-9720.2008.tb03297.x>
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46, 1–40. <http://dx.doi.org/10.1017/S0261444812000365>
- Matsuura, H. (2007). Intelligibility and individual learner differences in the EIL context. *System*, 35(3), 293–304. <http://dx.doi.org/10.1016/j.system.2007.03.003>
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25(3), 481–520. <http://dx.doi.org/10.2307/3586981>

- Munro, M. J. (2011). Intelligibility: Buzzword or buzzworthy? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 7–16). Ames, IA: Iowa State University. Retrieved from <http://jlevis.public.iastate.edu/2010%20Proceedings%2010-25-11%20-%20B.pdf>
- Neri, A., Cucchiari, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441–467. <http://dx.doi.org/10.1076/call.15.5.441.13473>
- Oda, K. (2012). Nihon no eigoka kyōiku ni okeru onsei shido no genjo [Research into phonetic teaching in English education at junior and senior high school in Japan]. *Institute Journal of Social Science and the Humanities*, 69, 53–73. Retrieved from <http://ci.nii.ac.jp>
- Ocker, R. J., & Yaverbaum, G. J. (1999). Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction. *Group Decision and Negotiation*, 8(5), 427–440. <http://dx.doi.org/10.1023/A:1008621827601>
- 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. <http://dx.doi.org/10.1111/tger.10181>
- Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46(4), 842–854. <http://dx.doi.org/10.1002/tesq.67>
- Schwienhorst, K. (2008). *Learner autonomy and CALL environments*. New York: Routledge.
- Suarez, S. & Tanaka, Y. (2001). Japanese learners' attitudes toward English pronunciation. *Niigata Seiryō Daigaku Kiyo [Bulletin of Niigata Seiryō University]*, 1, 99–111. Retrieved from <http://ci.nii.ac.jp>
- Szypra-Kozłowska, J. (2014). *Pronunciation in EFL instruction*. Bristol, UK: Multilingual Matters.
- Tominaga, Y. (2011). An analysis of English pronunciation of Japanese learners: From the viewpoint of EIL. *Journal of Pan-Pacific Association of Applied Linguistics*, 15(2), 45–57. Retrieved from <https://eric.ed.gov/?id=EJ979913>
- 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(1), 122–140. <http://dx.doi.org/10.1017/S1366728914000832>
- Tsurutani, C. (2008). A computer program for pronunciation training and assessment in Japanese language classrooms: Experimental use of 'Pronunciation Check'. *Japanese Studies*, 28(3), 305–315. <http://dx.doi.org/10.1080/10371390802446869>
- Tsurutani, C., Yamauchi, Y., Minematsu, N., Luo, D., Maruyama, K., & Hirose, K. (2006). Development of a program for self assessment of Japanese pronunciation by English learners. *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing* (pp. 841–844). International Speech Communication Association. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=0FD1CD444922BB94F85B5CC3120991C8?doi=10.1.1.418.6237>
- Version2. (2015). Glexa [LMS]. Sapporo: Version2
- Walker, R. (2005). Using student-produced recordings with monolingual groups to provide effective, individualized pronunciation practice. *TESOL Quarterly*, 39(3), 550–558. <http://dx.doi.org/10.2307/3588495>

- Ware, P.D., & O'Dowd, R. (2008). Peer feedback on language form in telecollaboration. *Language Learning & Technology*, 12(1), 43–63. Retrieved from <http://llt.msu.edu/vol12num1/wareodowd/>
- Wen, M. L., & Tsai, C.-C., (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, 51(1), 27–44. <http://dx.doi.org/10.1007/s10734-004-6375-8>
- Yoshida, K., Aliponga, J., Koshiyama, Y., Gamble, C., Wilkins, M., & Ando, S. (2012). Communicative language teaching in tertiary EFL contexts in Japan: Examining learner beliefs, perceptions and attitudes. In A. Stewart & N. Sonda (Eds.), *JALT2011 Conference Proceedings*, 120–134. Tokyo: JALT. Retrieved from http://JALT-publications.org/proceedings/issues/2012-10_2011.1
- Yonesaka, S. M., & Tanaka, H. (2014). *Discovering English sounds: Phonetics made easy*. Tokyo: Cengage Learning.
- Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36(1), 69–84. <http://dx.doi.org/10.1016/j.system.2007.11.004>

Author biodata

Suzanne M. Yonesaka is a professor in the Faculty of Humanities at Hokkai-Gakuen University, Sapporo, Japan.

Appendix

Pre-treatment Survey

Personal information

1. Sex: Male /Female
2. What is your native language (the language or languages that you grew up speaking)? _____
3. How many years have you studied English up until now?
6 years / 7 to 10 years
4. How long have you been in an English-speaking country?
Never / 1 to 30 days / 1 to 6 months / More than 6 months
5. How often do you speak in English for personal reasons? (e.g. with friends; at work)
Never / A few times each month / Several times a week / Every day

Pronunciation background

1. How would you assess your pronunciation of English sounds, compared to other students?
Extremely poor / Quite poor / Just OK / Pretty good / Good / Excellent
2. How would you assess your rhythm and intonation when you speak English, compared to other students?

Extremely poor / Quite poor / Just OK / Pretty good / Good / Excellent

3. What is your English pronunciation goal?
 I want to pronounce like a native-speaker. /
 I want to pronounce so that others will usually understand me.
4. Good pronunciation is important for communication so that the listener can understand the message correctly.
 Strongly agree / Agree / Slightly agree / Slightly disagree / Disagree /
 Strongly disagree
5. Have you ever received feedback (e.g., evaluation, opinion) on your pronunciation during class or outside of class?
 Yes / No
6. I feel comfortable getting feedback (e.g., evaluation, opinion, input) from other students about my pronunciation during class.
 Yes / No
7. It helps me when other students respond to my pronunciation during class.
 Yes / No

Post-treatment survey

Pronunciation background

(This is the same as Pre-treatment Survey, excluding question 5.)

P-Check

(Strongly agree / Agree / Slightly agree / Slightly disagree / Disagree / Strongly disagree)

1. I enjoyed the P-Check project.
2. Technological glitches with P-Check severely impeded the success of the project.
3. I became more conscious of my own pronunciation through this project (more so than I would have through just taking the phonetics course.)
4. Recording my own voice makes me nervous.
5. I learned a lot about my own pronunciation strengths and weaknesses through P-Check.