THE JOURNAL OF ASIA TEFL
Vol. 17, No. 3, Fall 2020, 824-840
http://dx.doi.org/10.18823/asiatefl.2020.17.3.5.824

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs

Solène Inceoglu
Australian National University, Australia

Hyojung Lim
Kwangwoon University, Korea

Wen-Hsin Chen
National Taipei University of Technology, Taiwan

The current study explored the usefulness of mobile-based automatic speech recognition (ASR) pronunciation practice by investigating a) its effects on the production of four English vowels, and b) learners' perception of ASR as a learning tool. A total of 19 Korean university students produced 28 minimal pair sentences containing the English vowel contrasts /i/-/ɪ/ and /ɛ/-/æ/ (e.g., I said beat, I said bit) at pretest and posttest, and completed six sessions of ASR practice outside of class that involved voice-typing a short text, minimal pairs in sentences, and decontextualized minimal pairs. Results of acoustic analysis of F1 and F2 formant frequencies showed a meaningful improvement in frontness for the vowel /i/, but no changes for the other vowels. Overall, the majority of the participants perceived ASR as useful for pronunciation practice, but some showed skepticism and frustration regarding the current state of the technology. Further discussed are the problems and limitations that EFL learners experienced during the ASR training.

Keywords: automatic speech recognition (ASR), EFL pronunciation, learners' beliefs, vowel production, pronunciation training

Introduction

The pedagogical benefits of automatic speech recognition (ASR), also referred to as speech-to-text, voice-recognition, or dictation, have been highlighted in second language (L2) teaching and learning research (Cucchiarini & Strik, 2018; Levis & Suvorov, 2012). The use of ASR appears to contribute to reading development in children (Mostow, Huang, & Junker, 2008) and ASR might be used to address common errors in L2 Dutch morphology and syntax (Cucchiarini, Van Doremalen, & Strik, 2008). In the domain of pragmatics, Chiu, Liou, and Yeh (2007) reported that ASR-based oral activities in a simulated real-life conversation helped Taiwanese freshman students learn English speech acts. Undeniably, the strongest contribution that ASR technology makes is in the area of oral skills and pronunciation. Among the affordances for language teaching and learning that ASR provides, Golonka et al. (2014, p. 73) listed that ASR can:

- "Compare student's pronunciation acoustically with a target pronunciation and provide feedback."
- "Provide learner with an opportunity to work on speaking ability individually, at self-selected pace."
- "Allow learner to practice simulated dialogue with computerized agent."

ASR is indeed particularly relevant for the development of learner autonomy (McCrocklin, 2016). ASR provides low-anxiety learning environments and immediate feedback which allows learners to practice their pronunciation in a repeated fashion. Not surprisingly, a number of commercial computer-assisted language learning (CALL) programs, such as Rosetta Stone and Tell me More, have incorporated ASR technology to encourage autonomous language learning outside of class, and numerous CALL researchers and speech technologists have worked on developing and testing ASR-based programs directed at a specific group of learners (e.g., Chiu et al., 2007; Cucchiarini, Neri, & Strik, 2009; Neri, Mich, Gerosa, & Giuliani, 2008; van Doremalen, Boves, Colpaert, Cucchiarini, & Strik, 2016). For instance, Van Doremalen et al. (2016) designed and evaluated a prototype of DISCO—an ASR-based language learning system, which provided feedback on the pronunciation, morphology, and syntax of L1 Dutch EFL learners. Their reviews showed that teachers and learners generally held a positive attitude towards the system and would consider using the language learning system in the future.

Outside the education context, the prevalence of non-commercial dictation programs on personal computers and ASR-based virtual personal assistant services (e.g., Apple's Siri, Google's Assistant, Amazon's Alexa, and SKT NUGU) opens up attractive new possibilities for language teachers and learners. Recently, attempts have been made to assess the pedagogical suitability of general (i.e., non-educational) ASR technology, focusing particularly on whether—and how well—these programs recognize non-native speakers' pronunciation and whether the feedback it provides is useful. McCrocklin, Humaidan, and Edalatishams (2019) reported that *Google Voice Typing* outperformed *Windows Speech Recognition (WSR)* in recognizing non-native production of sentence reading and free responses to open-ended questions. The authors suggested that *Google*'s ability to recognize non-native speech might be getting close to the level of accuracy of native listeners reported in the literature. However, *Google* showed a tendency to turn off while participants were dictating sentences, and both *Google* and *WSR* had higher levels of accuracy with native speech than with non-native speech.

In L2 speaking instruction, accuracy, alongside fluency, appropriacy, and authenticity, is one of the critical criteria that teachers and learners need to work on (Lazaraton, 2014). This is particularly the case in EFL settings where learners have limited opportunities to practice their oral skills. Yet, integrating pronunciation instruction in EFL curricula is often a challenge due to instructional and time constraints, and because non-native EFL teachers may lack pronunciation training and access to pronunciation teaching materials, resulting in low confidence (Darcy, 2018; Murphy, 2014). Pronunciation is also an aspect of the language that is associated with anxiety (Baran-Łucarz, 2011; Kermad, 2018; J. O. Kim, 2018). For instance, Korean EFL learners reported a higher level of anxiety both in perception and production at the segmental level (C. H. Kim & Kim, 2016). In this regard, ASR could be a promising tool to support teachers and help enhance EFL learners' pronunciation development, specifically through self-regulating practice.

The objectives of the current study were twofold: to explore the pedagogical benefits of autonomous mobile-based ASR training on Korean university EFL learners' vowel production, and to report on the learners' perception of the training. In particular, the study investigated whether ASR practice can lead to improvement in the production of two sets of English vowels /i/-/ɪ/ and /ɛ/-/æ/.

Literature Review

Second Language Learners' Perception of ASR

An increasing number of studies have corroborated the positive effect of ASR technology on learners' language development and learning experience. For instance, the use of ASR was found to increase learner awareness of their own pronunciation in French (Mroz, 2018), enhance motivation and enjoyment (Ahn & Lee, 2016; Guskaroska, 2019) , and increase learner autonomy (McCrocklin, 2016). McCrocklin (2016) showed that after a three-week ESL pronunciation workshop, both the hybrid group that received face-to-face instruction and *Windows Speech Recognition* practice, and the face-to-face group that received ASR strategy training, showed a significant increase in their beliefs in autonomy, whereas the non-ASR group did not. After the workshop though, only the hybrid group reported a significant increase in autonomous learning behavior and ASR pronunciation practice. Learners' positive attitudes towards ASR as a tool to practice speaking were also noted in other studies. For instance, Guskaroska's (2019) analysis of learners' Facebook posts revealed that the majority of the Macedonian EFL learners who participated in her study enjoyed ASR practice training sessions, and found the ASR tool useful and practical. Similarly, the middle school students in Ahn and Lee (2016) showed increased motivation for ASR-based speaking activities, primarily because of its interactive nature. Looking at what the future might bring for language learners both in and out of the classroom, Moussalli and Cardoso (2016) explored the perceptions of four intermediate learners of English about the use of Echo—a personal robot with built-in ASR software—as a pedagogical tool for learning English. Learners' responses showed that they enjoyed the experience and believed it to be useful for pronunciation and vocabulary learning. Echo was also found to provide helpful negative feedback that raised learners' awareness of problems or gaps in their production. In sum, learners' interaction with ASR technology has been reported to be positive, and ASR appears to be a potential means to promote learner autonomy in a stress-free environment. Yet, learners' engagement with ASR technology might also depend on learner beliefs and perceptions. Comparing learner usage (n = 2867) and perception (n = 482) of ASR built-in activities in blended-learning English courses, Artieda and Clements (2019) reported significant differences between countries of origin (Saudi Arabia, China, Vietnam, Italy) regarding learners' beliefs that technology can help them improve pronunciation, with Chinese participants scoring the lowest.

ASR and Pronunciation Learning

Empirical efforts have been made to capture the actual impact of computer-assisted pronunciation training (CAPT) and ASR on leaners' speaking ability. Wallace (2016) illustrated how *Google Web Speech* could help international teaching assistants in the United States notice their oral communication problems, which eventually contributed to mitigating their heavy accents and increasing the intelligibility of their production. Yet, the scope for improvement might also depend on the participants' pronunciation level at the beginning of the practice. Hincks (2003) compared 11 learners of English in Sweden who were given unlimited access to Talk to Me to a group of learners who did not use the ASR program. Although the overall effects of ASR were limited, she noted that using the program was beneficial for learners who had started the experiment with strong accented speech. In Wang and Young's (2015) study, adult and junior high school ESL learners participated in an 8-week pronunciation training where they practiced speaking English at their own pace in a web-based environment. The CAPT system provided formative and summative feedback in either a graphic, auditory, or textual form. In the pretest and posttest, participants were asked to listen to and repeat eight English sentences, and the accuracy of learner pronunciation was evaluated. Results showed that the adult learners improved their pronunciation more than the younger learners did, preferred auditory feedback—including the model pronunciation of full sentences and individual words, and found the summative feedback more beneficial for self-reflection.

While CALL research points to the potential of ASR as a learning tool for L2 speaking, empirical studies that have examined the actual effect of ASR on the development of L2 pronunciation, especially at the segmental level, are scant. Using a pre- and posttest design experiment, Liakin, Cardoso, and Liakina (2015) compared the production and perception of /y/ by three groups of beginning learners of French. One group performed ASR-based pronunciation exercises on a weekly basis, another group did the same activities but received teacher feedback, and a control group practiced conversation skills without any feedback. The results showed that although no group improved in perception, only the ASR group improved significantly in their production of /y/. In an EFL setting, Guskaroska (2019) investigated Macedonian learners' production of 30 words containing the minimal pairs /i/-/ɪ/, /æ/-/ɛ/, /u/-/ʊ/, and /ɑ/-/ʌ/. Learner pronunciation collected from the pretest and posttest was rated by 10 native speakers. Compared to the control group, the ASR group significantly improved their production accuracy for /u/, /æ/, and /ʌ/, but not for /i/ and /ɛ/. Similarly, McCrocklin (2019) compared one group of students who received face-to-face pronunciation training with a hybrid group who was given 50% of face-to-face training and 50% of ASR-based individual practice. Both groups participated in a three-week pronunciation workshop contextualized in a L2 English listening course in the U.S. The vowel contrasts (i.e., /ɛ/-/æ/, /ɑ/-/ʌ/, and /i/-/ɪ/) and consonants (i.e., /ɹ/, /θ/, /ð/, /ʒ/, and /dʒ/) were the target of instruction and subsequent analysis. Native speaker ratings revealed that both groups equally improved their pronunciation, thus showing that ASR can serve as a beneficial learning complement—especially when in-class time is limited. The learning effect, however, varied across vowels. Learners' production improved the most with the vowel /ɪ/ and /ɛ/ (about 10% accuracy gains), followed by /æ/ (about 5.1% accuracy gains), but retrogressed with the vowel /i/ (about 5.7% accuracy loss). Some of these results appear to contrast with Guskaroska's findings, which might be attributable to the differences in the participants' L1 and proficiency levels. While Guskaroska had a homogeneous group of Macedonian EFL students, McCrocklin looked at a mixed ESL group that included Chinese, Korean, Malay and Marathi learners of English. The amount and type of ASR training or rating methods could have affected the results as well. Finally, in our recent study with 49 Taiwanese EFL students (Chen, Inceoglu, & Lim, 2020), learners were better able to distinguish between /æ/and /ɛ/ in production after six sessions of ASR-based training, while showing no improvement with the /i/-/ɪ/ contrast.

In sum, the results of previous studies have highlighted that ASR practice can have beneficial effects on the development of learner autonomy and pronunciation learning, and that using ASR is usually perceived by learners as a positive, stress-free, and useful experience. Studies examining the effects of ASR pronunciation practice at the segmental level are however scant and limited to a very small group of L1/L2 participants. Accordingly, the goal of the current study was to further explore the usefulness of ASR dictation systems in the learning of English minimal pairs by Korean university EFL learners. The research questions guiding the present study were as below:

1) How does ASR pronunciation practice contribute to EFL learners' pronunciation development at the segmental level?
2) How do Korean EFL learners perceive their experience of using ASR to practice English pronunciation?

Methods

Participants

In spring 2019, 21 intermediate Korean EFL learners were recruited from a four-year university in Seoul, as part of their coursework or as volunteers. For data analysis, however, two students were excluded due to the low quality of their audio recordings. All the participants took a pre- and post-test and completed six ASR-based training sessions. Thirteen students were English majors taking an introduction

to computer-assisted language learning at the time of data collection, and six students from various majors were enrolled in a University English course. Three students reported having lived in an English-speaking country for three to four years. Only one student reported that she had used an ASR tool in English in the past for Internet searching; the rest did not use any ASR tool in English. Learner information collected from questionnaires is summarized in Table 1, including the level of language learning motivation and attitude towards pronunciation learning. At the start of the experiment, the learners were relatively neutral regarding how ASR could recognize their English and help them improve their pronunciation. Four native speakers of English (2 female) recorded the same word list that was provided to the L2 participants. Their vowel productions were analyzed and served as a baseline.

TABLE 1
Summary of Participants Characteristics

Age	21.47 (1.81)
Number	19 (8 female)
Major	English majors = 13, Others = 6
Age of formal English study	8.25 (1.86)
Length of stay in English-speaking countries (in years)	0.5 (1.35)
Exposure to English in a natural setting (e.g., watching YouTube in English) (in hours per week)	1.67 (1.58)
Pronunciation attitude score[1]	41.75 (3.48)
Belief that ASR can recognize their English pronunciation (before experiment)[2]	3.60 (1.14)
Belief that ASR practice can lead to pronunciation improvement (before experiment)[2]	3.65 (1.04)
Ideal L2 self score[2]	4.07 (1.34)
Ought-to L2 self score[2]	2.86 (1.53)

Note. Values given are means, with SDs in parentheses. [1] score out of 60. [2] scores from 1 "strongly disagree" to 6 "strongly agree."

Instruments and Procedures

This study is part of a larger project exploring the effects of ASR on EFL learners' pronunciation development in view of vowel quality, fluency, and intelligibility. Hence, the participants were asked to complete a wide range of speaking tasks, some of which were beyond the purpose of the current analysis and will not be described in detail.

Oral tests

All participants completed the pre- and post-test in the second author's office over a period from four to six weeks. The oral tests consisted of three types of speaking tasks: reading aloud a short passage, describing a set of pictures, and reading aloud a list of 14 minimal pairs. The target English vowel pairs were /i/-/ɪ/ and /ɛ/-/æ/ presented in full sentences, as in "I said bit, I said beat." It took about six to 10 minutes for the learners to complete the tasks. In this paper, we analyzed and reported the productions of six minimal pairs only (see Table 2), focusing on the pronunciation development in terms of the distinction between lax and tense English vowels.

TABLE 2
Pretest and Posttest Minimal Pairs Analyzed in this Study

	Target words embedded in sentence "I said X"	
/i/-/ɪ/	beat	bit
	sheep	ship
	peak	pick
/ɛ/-/æ/	said	sad
	dead	dad
	bed	bad

Questionnaires

In addition to the oral tests, the participants were asked to fill out two kinds of survey questionnaires: a pronunciation attitude inventory adapted from Elliott (1995) and a motivation questionnaire adapted from Dörnyei and Taguchi (2010). Each consisted of 16 items using a 6-point Likert scale with 6 being the highest (i.e., "strongly agree"). The pronunciation attitude inventory aimed to investigate the learners' belief in acquiring target-like pronunciation and included statements such as "I'd like to sound as native as possible when speaking English." The motivation questionnaire explored two dimensions of the learners' motivation proposed by Dörnyei (2009), 10 items for ideal L2 self (e.g., "Whenever I think of my future career, I imagine myself using English"), and six items for ought-to L2 self (e.g., "If I fail to learn this language, I'll be letting other people down"). All questionnaires were completed online immediately after the pretest. After the posttest, the participants completed an exit survey questionnaire, which targeted the learners' perception of the effectiveness of ASR-based pronunciation training and the usefulness of ASR as a language learning tool. The exit survey was made up of seven items using a 10-point Likert scale, in addition to two open-ended questions that prompted the participants to reflect on the most common problems they experienced recording themselves with ASR, and to elaborate on why they would or would not continue to use ASR to practice their English pronunciation.

Pronunciation training materials

All students participated in six sessions of ASR practice outside of class over a period of three weeks. The practice set was similar to the oral pretest and posttest, and consisted of three parts: reading aloud a short reading passage (average length of 104 words), reading short sentences with minimal pairs as in "I took a sip of water on the ship" (four sentences per session), and reading decontextualized minimal pairs (four minimal pairs per session). The feedback the participants received on their pronunciation was not explicit but was available through the written output provided by the ASR program. That is, the students looked at their screens to check whether the ASR written output matched their intended spoken output (i.e., the sentences and words they aimed to produce). Each session presented different texts, sentences, and word lists, and took approximately three to six minutes per session, depending on the students' engagement with the tasks. The participants were asked to video-record each training session with a screencast tool that captured the ASR written output and the participants' voices and email it to their instructor. Upon receiving the video file, the instructor sent the next practice set to each student. Given that an important goal of computer-assisted pronunciation learning is to foster learner autonomy (McCrocklin, 2016), the participants were allowed to work on the practice set at their own pace. We did not impose the type of ASR program that students adopted for practice; eight participants used Google's Assistant available for Android, 10 used Apple's Siri, and one mixed both Google's Assistant and Siri.

Data Analysis

For the current analysis, we ran acoustic analyses on the learners' vowel productions to investigate how ASR-based pronunciation training impacted the development of two sets of vowel contrasts. A total of 504 tokens were collected from the pretest and posttest recordings. Following Boersma and Weenink (2018), we manually extracted the F1 and F2 measurements from each target word at mid-point, in addition to vowel duration. The first formant (F1) correlates inversely with tongue height and the second formant (F2) correlates inversely with tongue backness; in our current study, for instance, /i/ is the highest and most fronted vowel. The analysis of four L1 speakers (2 female) who recorded the same stimuli served as a reference point and is reported in Table 3.

TABLE 3
Mean Acoustic Values of the English Native Speakers' Vowels

	/i/	/ɪ/	/ɛ/	/æ/
F1	316 (39)	464 (76)	620 (89)	755 (86)
F2	2494 (347)	2034 (163)	1858 (176)	1717(296)
Duration	103 (3)	81 (3)	124 (4)	105 (4)

Note. Values given are means, with SDs in parentheses.

Vowel durations and F1 and F2 measurements of the pretest and posttest were submitted to paired *t*-tests to investigate the development of learners' vowel pronunciation. Because four *t*-tests were performed, a Bonferroni adjustment required an alpha level of .0125 (.05/4) to be used for all statistical tests reported below. Learners' responses to the questionnaires were analyzed based on the frequency count and the qualitative comments.

Results and Discussion

Pronunciation Development at the Segmental Level

To answer the first research question, we analyzed the acoustic features of the vowels that the learners produced in the contextualized minimal pairs. Paired *t*-tests revealed that ASR training led to significant yet marginal changes to the quality of learners' vowel production. However, the patterns were not consistent across vowels. For /i/ (Table 4), results showed no significant changes in F1 values, $t(125) = 1.448$, $p = .150$, but a significant increase in duration, $t(125) = -8.238$, $p < .001$, and in F2 values, $t(125) = -4.590$, $p < .001$, with L2 learners approaching native speakers' frontness at posttest.

TABLE 4
The t-Test Results of /i/

		Mean	*N*	*SD*	*SE*
Duration ($p < .001$)	Pretest	162	126	49.5	4.42
	Posttest	197	126	67.2	5.99
F1 ($p = .150$)	Pretest	382	126	77.5	6.91
	Posttest	375	126	63.6	5.66
F2 ($p < .001$)	Pretest	2355	126	425.2	37.88
	Posttest	2454	126	428.2	38.15

For /ɪ/ (Table 5), there was a significant increase in duration, $t(125) = -3.305$, $p < .001$, but no significant changes either in F1 values, $t(125) = -1.026$, $p = .307$, or in F2 values, $t(125) = .049$, $p = .961$. The data revealed that our Korean EFL learners did not distinguish /i/ and /ɪ/ in production neither at pre-

nor post-test, thus indicating that six sessions of ASR training did not lead to improved pronunciation for this vowel. One way of accounting for this lack of change is to suggest that ASR might provide some type of feedback, but does not involve a model (i.e., native speakers' production of the vowel) without which learners are not able to modify their pronunciation. Because speech perception and production are linked (Best, 1995; Flege, 1995), ASR practice could benefit from being paired up with high-variability perceptual training (Thomson, 2011) for L2 segments that are particularly challenging.

TABLE 5
The t-Test Results of /ɪ/

		Mean	*N*	*SD*	*SE*
Duration	Pretest	114	126	39.0	3.47
	Posttest	125	126	39.5	3.52
F1	Pretest	420	126	85.1	7.58
	Posttest	426	126	81.9	7.30
F2	Pretest	2166	126	420.9	37.50
	Posttest	2164	126	382.1	34.04

The results for /ɛ/ revealed a significant increase in F2 values, $t(125) = -5.69$, $p < .001$, and in duration, $t(125) = -5.632$, $p < .001$, but these changes were not positive as they moved away from the native speaker norms. On the other hand, there were no significant changes in F1 values, $t(125) = .249$, $p = .804$ (Table 6) and the learners' vowel height remained at about 200 Hertz, too low compared to the native speakers' norms.

TABLE 6
The t-Test Results of /ɛ/

		Mean	*N*	*SD*	*SE*
Duration	Pretest	192	126	58.0	5.17
	Posttest	219	126	78.4	6.99
F1	Pretest	682	126	131.6	11.73
	Posttest	680	126	146.9	13.09
F2	Pretest	1828	126	235.4	20.98
	Posttest	1888	126	239.0	21.29

Finally, for /æ/ (Table 7), there were significant increases in the wrong direction in F2 values, $t(125) = -2.285$, $p = .024$, and in duration, $t(125) = -4.491$, $p < .001$, with /æ/ being produced too long and too fronted compared to the native speaker baseline. Despite a positive trend in the F1 change, the results were not statistically significant, $t(125) = -1.743$, $p = .084$, with the vowel remaining too high.

TABLE 7
The t-Test Results of /æ/

		Mean	*N*	*SD*	*SE*
Duration	Pretest	186	126	61.7	5.50
	Posttest	209	126	65.2	5.80
F1	Pretest	691	126	133.3	11.88
	Posttest	703	126	139.6	12.44
F2	Pretest	1803	126	255	22.72
	Posttest	1846	126	277.8	24.75

The longer vowel duration observed during the posttest could be a simple indicator that the learners made extra efforts to articulate each word. Although beyond the scope of the current study, observations of the ASR practice videos reveal that the learners tended to exaggerate vowel length in attempts to have

their pronunciation recognized by the ASR system. This also adds to Korean EFL learners' general tendency to produce longer vowel duration (Lee & Cho, 2013).

Figure 1 illustrates the participants' average vowel space at pretest and posttest, along with a reference to native speakers' average norms. The data shows that Korean EFL learners distinguish /ɪ/ from /i/, but to a much lesser extent than native speakers. This is consistent with Yang's (2013) study, in which Korean participants showed a shorter distance of the front vowel pair than American counterparts. In addition, as shown in the acoustic analyses above, the Euclidean distance between each vowel did not become larger from pretest to posttest. On the other hand, the learners tended to group /ɛ/ and /æ/ into one vowel space, with very little change from pretest to posttest. This is in line with a number of studies highlighting Korean learners' difficulty to distinguish between /ɛ/ and /æ/ in both perception and production (e.g., Ingram & Park, 1997; Kim, 2010; Yang, 2013). As such, the current data may suggest that ASR pronunciation training has varying degrees of effectiveness across vowels and is not sufficient for vowel pairs that are very difficult.

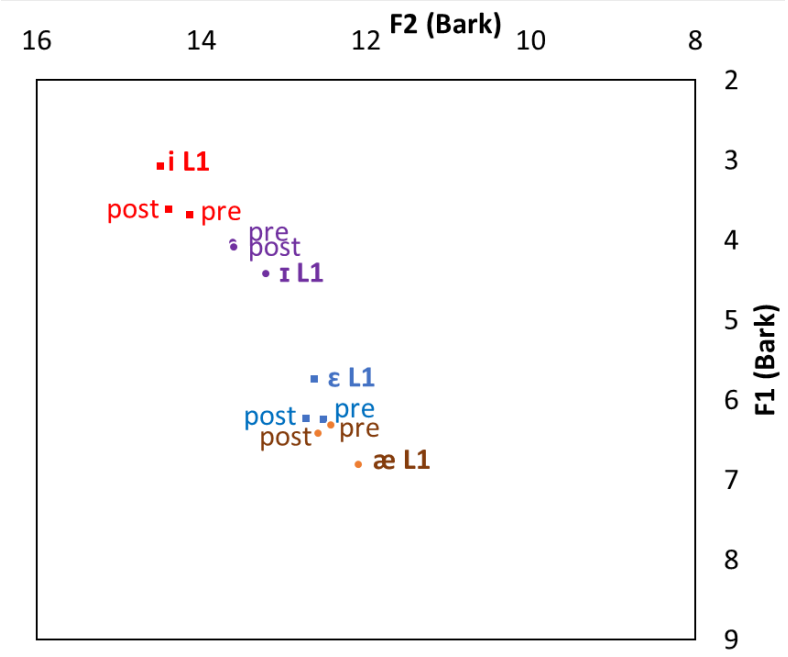

Figure 1. Average formant values of /i/, /ɪ/, /ɛ/ and /æ/ by Korean EFL participants at pretest and posttest (L1 represents the native English norms).

Awareness Raising and Students' Beliefs

Overall, the students were generally positive in their evaluation of ASR for the purpose of practicing English pronunciation in texts, sentences, and minimal pairs. In terms of the perceived usefulness of ASR, 13 out of 19 students (68%) reported finding ASR helpful to practice English pronunciation, two students (10.5%) remained neutral, and four students (21%) rated its usefulness less than the midpoint (Figure 2).

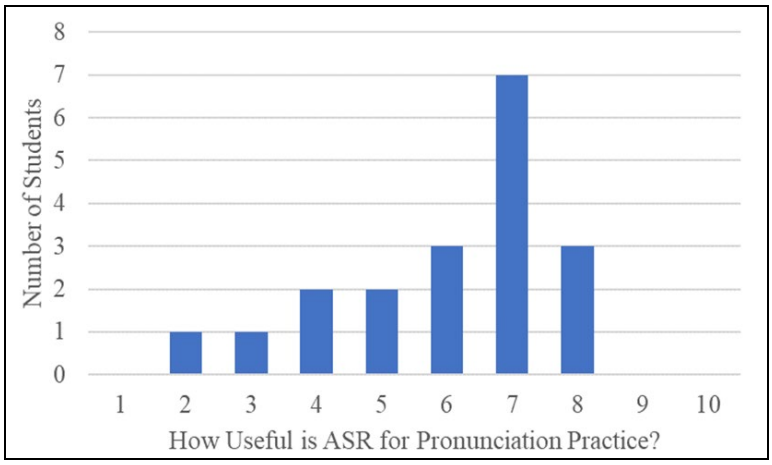

Figure 2. Students' ratings of the usefulness of ASR as a pronunciation practice tool (from 1 "terrible" to 10 "excellent").

When asked whether they intended to use ASR as a pronunciation practice tool in the future, however, 58% of the students (n = 11) noted that they would continue to use the ASR tool for pronunciation practice, while 42% of the students (n = 8) said that they would not. Among the positive comments, the students emphasized the beneficial impacts of the immediate feedback that ASR provides. Because EFL learners are often faced with limited opportunities to practice their language and receive feedback on their pronunciation, ASR was considered by some learners in the study as a useful pronunciation *learning* tool—or to the very least as a useful pronunciation *practice* tool. The examples of students' quotes below were translated from Korean.

> ASR is great because it provides very objective feedback. It is also easy to access. So, I would use it more frequently to check out my pronunciation. (ID 15)

> The ASR program elicits repetitions. The repetition itself helps with learning. (ID 12)

> I think ASR recognizes my pronunciation fairly well. Why not using it? (ID11)

> I would use it to check out the accuracy of my pronunciation. However, I do not feel that the use of the ASR tool contributes to my pronunciation development to a great extent. (ID 9)

One learner demonstrated not only excellent pronunciation learning strategies, but also a conscious effort to improve his English, and commented on the value of ASR as a vocabulary learning tool, especially to strengthen the relationship between orthography, phonology, and meaning, and create links between pronunciation learning and other areas of L2 learning, including vocabulary acquisition (Sicola & Darcy, 2015):

> I look up a dictionary to double-check the pronunciation of the words that ASR misunderstands, which actually helps me much with vocabulary learning. So, I will continue to use the ASR tool for language learning. (ID 17)

On the other hand, the learners who reported no intention of using ASR for future learning emphasized some limitations of ASR as a learning tool. In particular, half of the negative group (n = 4) cast doubt on the quality of ASR technology, its accuracy, and thus the usefulness of its feedback.

If the quality of ASR improves further, then I will consider using it for language learning. (ID 13)

I found ASR somewhat limited in accurately recognizing my pronunciation. (ID 5)

Additionally, one student wished to have a model of native speaker pronunciation to follow, which is available in other online tools (e.g., online dictionaries):

For pronunciation development, other options work better than ASR, such as online dictionaries or YouTube lectures on pronunciation. (ID 20)

Two students criticized the fact that ASR does not provide "explicit feedback," which is an explanation of why their pronunciation was incorrect. However, it seems unclear to what extent ASR feedback should be considered explicit. Given that it detects errors at the word or even phoneme level, ASR feedback could be said to be relatively explicit (Cucchiarini & Strik, 2018). However, in the absence of detailed pedagogical explanations (i.e., what in the pronunciation of a word is erroneous) or comparison with native-like models, the feedback dictation-based ASR programs give is limited and tends to elicit discovery learning, all of which could account for students perceiving the feedback as indirect or not explicit. Presumably, the students below might not understand what their pronunciation errors were and this failure to self-assess and monitor themselves led to confusion and frustration:

It doesn't provide explicit corrective feedback. So, I never know why my pronunciation doesn't work. (ID 5)
It might have been better if ASR could pinpoint which part of my pronunciation was wrong. (ID 13)

Finally, two students appeared to consider ASR purely as a way to voice-type messages instead of as a tool to practice pronunciation, thus highlighting a gap between the pedagogical nature of the task and the students' perception of the task:

It's slower than typing, so I don't think I'll use it often. (ID 7)

I do not use voice-typing on a daily basis. So, I don't feel comfortable with using ASR. (ID 8)

Regarding how well ASR was judged to recognize learners' pronunciation, 13 students reported that ASR picked out their pronunciation properly, rating higher than the midpoint (Figure 3). In cases where ASR failed to understand learners' pronunciation, only one participant viewed his pronunciation as the sole source of the problem, while most (n = 13) attributed the failure to both the low quality of ASR technique and their pronunciation. This finding can be linked to past research showing that learners' self-assessment of their speech might not be very accurate, with learners at the low end of the accentedness and comprehensibility scale mostly overestimating their performance (Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2016). In the context of ASR pronunciation practice, some students might have come to believe that their pronunciation was accurate but not recognized by ASR, when in fact they might not have been as intelligible as they believe to be.

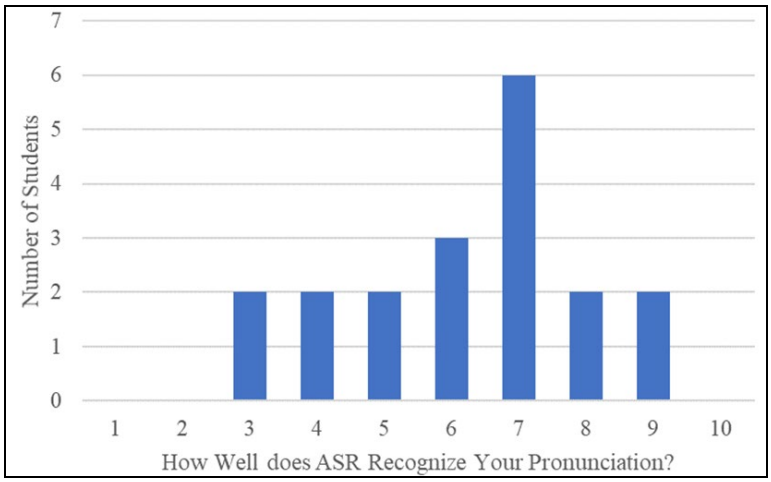

Figure 3. Students' ratings of how well ASR recognizes their speech (from 1 "terrible" to 10 "excellent").

The posttest survey also asked the participants to report the most common problems they encountered during the ASR training sessions. Several themes emerged from the participants' free responses. Below, we present the main issues along with examples of the participants' quotes, starting with the most frequent problems—based on how many students raised that point—to less frequently mentioned problems.

1. The main issue reported by the students was that ASR failed to understand their pronunciation, especially when it came to minimal pairs (n = 9):

   ASR does not get my minimal pairs. (ID 19)

   ASR never understands some of my pronunciation. (ID 7)

   It [ASR] does not understand the difference between [r] and [l], and between sit, seat, sheet, and shit. (ID 15)

   It [ASR] has problems recognizing similar pronunciation such as sit and seat. (ID 14)

2. A second issue highlighted by the students was linked to ASR's automatic correction, that is, instances when ASR provides a word before automatically changing it, sometimes due to the sentence context (n = 6):

   I pronounced a word correctly, but ASR automatically changed the form by using the context. Very frustrating. (ID 9)

   ASR automatically changes the form, which I did not intend. (ID 18)

   When practicing similar words (speaking out the minimal pairs in a consecutive manner), pronunciation itself was difficult and the ASR's automatic correction based on the context makes the practice difficult. (ID 6)

3. The third most common comment emphasized the students' beliefs that ASR technology is not advanced or reliable enough, and therefore did not seem to help with pronunciation practice (n = 3)

ASR does not seem sensitive to speakers' speaking rate and stress patterns. (ID 8)
Though I made efforts to make the sounds different, ASR still perceived them identical. (ID 14)

ASR built in Naver and Siri sometimes stopped working. (ID 16)

In addition to the major issues mentioned above, some students reflected that using ASR for pronunciation practice was time-consuming. Besides the likely underestimation of how much time and effort is needed to learn a foreign language, a possible reason behind the "time issue" is that the constant automatic changes in the written output displayed by the software algorithm and the inability of ASR to understand the learners' pronunciation—both of which were frequently mentioned in learner responses, contributed to frustration and the perceived inefficiency of the learning process. Importantly, only one student reported having encountered no problem using ASR.

General Discussion and Conclusion

To date, the few studies that have investigated ASR practice on segmental learning have shown that although ASR can be beneficial for pronunciation improvement, the effects depend on the target L2 sounds and the participants' L1 (Chen et al., 2020; Guskaroska, 2019; Liakin et al., 2015; McCrocklin, 2019). The current study adds support to this claim by providing evidence that ASR dictation practice was beneficial for the /i/-/ɪ/ contrast, but not for /ɛ/-/æ/. Interestingly, our previous study with Taiwanese EFL learners showed the opposite pattern, namely that after six sessions of autonomous ASR training, with the exact same materials presented in this study, Taiwanese learners moderately improved in their production of /ɛ/-/æ/, but continued to produce /i/-/ɪ/ in a single non-nativelike category (Chen et al., 2020). These mixed effect results have also been reported by Guskaroska (2019) whose Macedonian EFL learners improved in their production of /æ/ (but not /ɛ/ and /i/), and McCrocklin (2019) who found a high level of improvement for /ɪ/ but a decrease in accuracy for /i/.

The duration of the ASR practice is also a factor to consider. Studies so far, including the present one, have limited their focus to a two- to three-week intervention, and a longer duration—preferably over a semester—might have led to stronger improvement. Considering that ASR is a practical tool that can be used outside of class, further studies should investigate its effects, along with students' developing strategies, in a longitudinal way.

The second goal of this study was to investigate students' perception of ASR as a tool to practice pronunciation. Overall, and in line with previous studies (Guskaroska, 2019; McCrocklin, 2016), the majority of the students reported finding ASR helpful. Considering that when initially asked how much they thought ASR could lead to pronunciation improvement, the students were neutral (i.e., mean score of 3.65 out of 6), the results of the exit survey highlight a positive first experience. Also positive is the fact that some students reported using additional resources, such as online dictionaries, to check the pronunciation of words while doing their ASR practices. This demonstrates that ASR not only served as a tool to practice pronunciation, but also contributed to autonomous learning. It is also worth mentioning that the participants in the current study reported positive attitudes towards pronunciation, in addition to strong ideal L2 selves which have been shown to be strongly associated with one's desire to improve pronunciation (Huensch & Thompson, 2017). Although independent from the ASR practice, these results are promising as they highlight participants' interest in L2 pronunciation learning.

Nevertheless, not all students' feedback regarding ASR was positive. Some learners either did not find ASR practice meaningful, struggled with the task, or failed to see a connection between ASR and pronunciation learning. From a pedagogical point of view, this suggests that more guidelines could have been given to students regarding how to make the most use of ASR to practice speaking, thus maximizing student agency (Blake, 2013). For instance, integrating even just a short session of ASR practice at the beginning of the training, providing screen-recordings of native speakers dictating sentences on their

phones, combining ASR practice with shadowing activities (Hamada, 2016), or discussing both pronunciation and ASR technical strategies for the tasks might have enhanced the experience of some of the learners who reported negative views of ASR as a pedagogical tool. This echoes Romeo and Hubbard's (2010) CALL principles of "learner training," whereby learners should receive *technical training*—how to use the technology, *strategic training*—what to do to enhance learning, and *pedagogical training*—understanding why certain techniques and procedures are used.

While the current study provided insights into the effectiveness of ASR dictation practice on segmental accuracy and on EFL learners' perception of ASR as a pedagogical tool, there are some limitations that need to be addressed. First, not only was the number of participants relatively small, the number of items used for the analysis was also limited. In particular, the present study looked at only two minimal pairs of English vowels in a small set of words. Second, the analysis only focused on the participants' pretest and posttest oral production; the accuracy of feedback provided during the ASR training sessions was beyond the scope of the study. It would be interesting for future studies to focus on the mobile-phone screen recordings—capturing both the ASR output and the participants' voices—to examine how accurate mobile-based ASR technology is at providing feedback. For instance, do learners who mispronounce words receive false positive feedback or, conversely, does ASR provide erroneous written output even when the pronunciation is accurate? Future analyses should also examine the contribution of individual differences, especially with regards to the relationship between pronunciation development, motivation, and attitude towards ASR. The data presented in the current study captured high variability in learners' performance and perception of ASR and more research is needed on this topic.

Acknowledgements

The research was supported by the 2020 Kwangwoon University Research Fund.

The Authors

Solène Inceoglu is Lecturer (Assistant Professor) in the School of Literature, Languages and Linguistics at the Australian National University. She received her Ph.D. in Second Language Studies from Michigan State University. Her research focuses on second language acquisition, second language speech perception/production, pronunciation instruction, and psycholinguistics.

School of Literature, Languages, and Linguistics
Australian National University
Baldessin Precinct Building,
Acton, ACT 2601, Australia
Tel: + 61 2 6125 3532
Email: solene.inceoglu@anu.edu.au

Hyojung Lim (corresponding author) is an assistant professor in the Department of English language and Industry at Kwangwoon University, Korea. Her research interests revolve around second language vocabulary acquisition, computer-assisted language learning, and language testing.

Department of English Language and Industry
Kwangwoon University
Nowongu 01879, Seoul, Korea
Tel: +82-2-940-8322
Email: lim@kw.ac.kr

Wen-Hsin Chen is an assistant professor in the Media Center at National Taipei University of Technology, Taipei, Taiwan. Her research interests include second language acquisition, language classroom interaction, language processing, and Chinese language and culture.

Media Center, National Taipei University of Technology
No. 1, Sec. 3, Zhongxiao E. Rd., Da-an Dist.,
Taipei City 10608, Taiwan (R.O.C.)
Tel: +886-2-2771-2171 ext. 1144
Email: wchen33@ntut.edu.tw

References

Ahn, T., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*, 778–786.

Artieda, G., & Clements, B. (2019). A comparison of learner characteristics, beliefs, and usage of ASR-CALL systems. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouësny (Eds.), *CALL and complexity: Short papers from EUROCALL 2019* (pp. 19–25).

Baran-Łucarz, M. (2011). The relationship between language anxiety and the actual and perceived levels of foreign language pronunciation. *Studies in Second Language Learning and Teaching*, *1*, 491–514.

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

Blake, R. J. (2013). *Brave new digital classrooms: Technology and foreign-language learning* (2nd ed.). Washington D.C: Georgetown University Press.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer (Version 6.0.42) [Computer software]. Retrieved from http://www.praat.org.

Chen, W., Inceoglu, S., & Lim, H. (2020). Using ASR to improve Taiwanese EFL learners' pronunciation: Learning outcomes and learners' perceptions. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference* (pp. 37–48). Ames, IA: Iowa State University.

Chiu, T. L., Liou, H. C., & Yeh, Y. (2007). A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning*, *20*, 209–233.

Cucchiarini, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, *51*, 853–863.

Cucchiarini, C., & Strik, H. (2018). Automatic speech recognition for second language pronunciation training. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation.* (pp. 556–569). London: Routledge.

Cucchiarini, C., Van Doremalen, J., & Strik, H. (2008). DISCO: Development and integration of speech technology into courseware for language learning. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2791–2794.

Darcy, I. (2018). Powerful and effective pronunciation instruction: How can we achieve it? *The Catesol Journal*, *30*, 13–45.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford, UK: Oxford University Press.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language tesearch: Construction, administration and processing* (2nd ed.). New York, NY: Routledge.

Elliott, A. R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal*, *79*, 530–542.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*, 70–105.

Guskaroska, A. (2019). *ASR as a tool for providing feedback for vowel pronunciation practice* (Unpublished master's thesis). Iowa State University, Ames.

Hamada, Y. (2018). Shadowing for pronunciation development: Haptic-shadowing and IPA-shadowing. *The Journal of Asia TEFL*, *15*(1), 167–183.

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL*, *15*, 3–20.

Huensch, A., & Thompson, A. S. (2017). Contextualizing attitudes toward pronunciation: Foreign language learners in the United States. *Foreign Language Annals, 50,* 410–432.

Ingram, J. C. L., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343–370.

Kermad, A. (2018). *Speaker and listener variability in the perception of non-native speech* (Unpublished doctoral dissertation). Northern Arizona University.

Kim, C. H., & Kim, H. Y. (2016). EFL Learners perceptions of pronunciation, pronunciation anxiety, and pronunciation learning strategies. *The Journal of English Language and Literature*, *21*, 265–284.

Kim, J.-E. (2010). Perception and production of English front vowels by Korean speakers. *Phonetics and Speech Sciences*, *2*, 51–58.

Kim, J.-O. (2018). Ongoing speaking anxiety of Korean EFL learners: Case study of a TOEIC intensive program. *The Journal of Asia TEFL, 15*(1), 17–31.

Lazaraton, A. (2014). Second language speaking. *Teaching English as a Second or Foreign Language*, *4*, 106-120.

Lee, K.-Y., & Cho, M.-H. (2013). Production of English vowels by Korean learners. *The Journal of the Korea Contents Association*, *13*, 495–503.

Levis, J., & Suvorov, R. (2012). Automatic speech recognition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 316–323). New York, NY: Wiley-Blackwell.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*, 1–25.

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25–42.

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, *5*, 98–118.

McCrocklin, S., Humaidan, A., & Edalatishams, I. (2019). ASR dictation program accuracy: Have current programs improved? In *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 191–200). Ames, IA: Iowa State University.

Mostow, J., Huang, C., & Junker, B. (2008). 4-month evaluation of a learner-controlled reading tutor that listens. In V. M. Holland & F. P. Fisher (Eds.) *The path of speech technologies in computer assisted language learning* (pp. 215–233). London: Routledge.

Moussalli, S., & Cardoso, W. (2016). Are commercial 'personal robots' ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture: Short papers from EUROCALL 2016* (pp. 325–329).

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*, 617–637.

Murphy, J. M. (2014). Teacher training programs provide adequate preparation in how to teach pronunciation. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 188–224). Ann Arbor, MI: University of Michigan Press.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*, 393–408.

Romeo, K., & Hubbard, P. (2010). 15 pervasive CALL learner training for improving listening proficiency. In M. Levy, F. Blin, C. Bradin Siskin, & O. Takeuchi (Eds.), *WorldCALL, International perspectives on computer-assisted language learning* (pp. 215–229). London: Routledge.

Sicola, L., & Darcy, I. (2015). Integrating pronunciation into the language classroom. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 471–487). Malden, MA: Wiley-Blackwell.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*, 744–765.

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*, 122–140.

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, *29*, 833–851.

Wallace, L. (2016). Using Google web speech as a springboard for identifying personal pronunciation problems. In *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, 180–186.

Wang, Y. H., & Young, S. S. C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, *31*, 493–504.

Yang, B. (2013). A comparative study of relative distances among English front vowels produced by Korean and American speakers. *Phonetics and Speech Sciences, 5*, 99–107.

(Received May 20, 2020; Revised August 30, 2020; Accepted September 09, 2020)