English Teaching, Vol. 78, No. 1, Spring 2023, pp. 3-23
DOI: https://doi.org/10.15858/engtea.78.1.202303.3
http://journal.kate.or.kr

Effects of Pronunciation Training Using Automatic Speech Recognition on Pronunciation Accuracy of Korean English Language Learners

Thomas Dillon and Donald Wells*

Dillon, Thomas, & Wells, Donald. (2023). Effects of pronunciation training using automatic speech recognition on pronunciation accuracy of Korean English language learners. *English Teaching, 78*(1), 3-23.

This study examined effects of pronunciation training using automatic speech recognition technology on common pronunciation errors of Korean English learners. Participants were divided into two groups. One group was given instruction and training about the use of automatic speech recognition for pronunciation practice. The other group was not given such instruction or training as a control group. A pre- and post-test experimental design was used. The treatment period was four weeks. Participants who were taught about using automatic speech recognition for pronunciation practice showed small but significant improvements in pronunciation accuracy than those who did not. In addition, automatic speech recognition was found to assist in the diagnostic evaluation of common pronunciation errors, although it did not produce statistically significant improvements. Participants responded positively to the use of automatic speech recognition for pronunciation practice and testing, although there remain some concerns over technical aspects of the test.

Key words: automatic speech recognition (ASR), mobile assisted language learning, pronunciation practice, word error, EFL learner

This work was supported by Daegu Catholic University.

*First Author: Thomas Dillon, Professor, Foreign Language Education Centre, Daegu Catholic University
Corresponding Author: Donald Wells, Professor, Foreign Language Education Centre, Daegu Catholic University; 13-13, Hayang-ro, Hayang-eup, Gyeongsan-si, Gyeongsangbuk-do, 38430, Korea; Email: dillon@cu.ac.kr

Received 31 December 2022; Reviewed 30 January 2023; Accepted 17 March 2023

1. INTRODUCTION

With the improvement of voice recognition technology, there has been increased attention toward the use of automatic speech recognition (ASR) in the English language classroom. While research on the use of Computer Assisted Pronunciation Training (CAPT) is well established, ASR has emerged more recently as a viable technological option (Hsu, 2016). Advancements in natural language processing now permit commercially-available ASR products to output intelligible results even to non-expert users (Daniels & Iwago, 2017). It is now possible for teachers and learners to utilize ASR from their smartphones.

It is widely accepted that the development of a learner's pronunciation skill is essential to English language learning; however, in many English as a Foreign Language (EFL) or English as a Second Language (ESL) contexts, teachers often find themselves with insufficient resources to teach pronunciation effectively. Problems range from large class sizes with widely varying English proficiency levels (Chen & Goh, 2011), to textbooks which are ineffective at integrating pronunciation and speaking (Hayati, 2010).

Given the increasing capacity and availability of ASR, together with the practical limitations of EFL and ESL classrooms, researchers have begun to investigate the effectiveness of this technology in improving the pronunciation of second language (L2) learners. Some studies (e.g., Golonka, Bowles, Frank, Richardson, & Freynik, 2014; McCrocklin, 2019; Spring & Tabuchi, 2022) have suggested that such technology can be effective in producing meaningful improvements in pronunciation; however, there have been few studies that provided statistical evidence of improvements. The present study is an effort to identify the benefits ASR may provide with overall pronunciation accuracy as well as with specific errors, as compared with ordinary classroom practice not utilizing this technology.

The current study aims to build on previous research by answering the following three questions:

1) Does guided ASR practice improve Korean L2 English learners' overall pronunciation accuracy?
2) Does guided ASR practice help Korean L2 English learners avoid common pronunciation errors?
3) How do students evaluate guided ASR practice?

2. LITERATURE REVIEW

2.1. Pronunciation in Language Learning

While it is accepted that learning pronunciation is essential to second language (L2) speaking, it is a matter of controversy as to what constitutes an acceptable level of pronunciation mastery for a learner. Previous approaches which focused on achieving a native-like accent were criticized by, e.g., Holliday (2006) as "native speakerism." Following such critiques, researchers such as Levis (2018) and Munro (2011) have argued for achieving an intelligible accent, that is, an accent that can be understood by others, as a reasonable goal. However, this goal remains difficult to achieve for some speakers, particularly those without access to immersion in an L2 environment or without opportunities for significant L2 interaction (Baker & Burri, 2016).

What constitutes an intelligible accent can be difficult to measure. Moreover, teachers often lack the necessary resources to properly evaluate and train the pronunciation skills of individual students. For example, in teaching environments such as the 50-student Chinese classroom discussed by Liu et al. (2019), offering feedback to individual students proved overwhelmingly difficult. Even in cases where native speakers have the time and ability to assess intelligibility directly, human errors can taint the evaluation process (Lindemann & Subtirelu, 2013). The use of automatic speech recognition (ASR) may provide a solution to both of these problems by offering scalable and objective assessments of pronunciation. This possibility is supported by the observation that accurate ASR transcription of L2 speakers' pronunciation correlates with native speakers perceiving L2 speakers' pronunciation as highly accurate (Ashwell & Elam, 2017; Mroz, 2018).

2.2. The Potential Effectiveness of Automatic Speech Recognition

Given the potential usefulness of ASR, researchers have increasingly turned their attention to its utility in the classroom. Studies have generally found positive responses to the use of ASR-based technology in the classroom. Ahn and Lee (2016) found Korean middle school students were very receptive to the use of ASR, particularly its capacity to provide immediate feedback and to provide an interactive activity. A study of Taiwanese adult learners likewise found a high level of satisfaction with the use of ASR (Evers & Chen, 2022). A prior study by the present researchers also recorded a high level of appreciation of and engagement with ASR technology among Korean university students (Dillon & Wells, 2021).

A growing body of research suggests that practice with ASR can bring modest benefits to pronunciation skills. McCrocklin (2019) found that ASR use could provide students with additional resources and tools to monitor their progress and receive feedback on their errors. Inceoglu, Lim and Chen (2020), in a study of Korean speakers, found modest improvement in pronunciation through the use of ASR, albeit with some skepticism and frustration among participants with the current state of ASR technology.

More promising is the meta-study of Golonka et al. (2014). This study reviewed the findings of over 350 prior studies of various technologies used in the classroom and highlighted ASR in particular for its potential value in improving pronunciation and providing effective feedback to learners. Xiao and Park (2021) found that ASR use could provide benefits to users with a variety of pronunciation issues and also that it could accurately diagnose errors. Spring and Tabuchi (2022) found that ASR-based pronunciation practice conducted over the course of a semester at a university in Japan had a small but positive impact on pronunciation. The small but positive impact found by Spring and Tabuchi is in line with similar results found by Dai and Wu (2021) and Evers and Chen (2022).

2.3. Data Collection and Automatic Speech Recognition

Pronunciation is often assessed using rubrics, but this approach has been criticized by a number of researchers (e.g., Ghoorchaei & Rahmani, 2018; Levis, 2006). Various approaches to the creation of rubrics have been proposed to deal with the criticisms (e.g., Dai & Wu, 2021; Penkhae, 2020; Yulia, 2018). Yulia (2018) created a rubric for performance assessment with nine dimensions and a three point scale which included characterization and non-verbal communication. Penkhae (2020) focused on difficult sounds using a subjectively rated rubric with a 5 point scale. Dai and Wu (2021) used a 3-dimension rubric with a binary scale, consisting of comprehensibility, segmental accuracy and word stress accuracy. These rubrics had several dimensions (e.g., stress, rhythm, intonation) that could not be diagnosed using a speech-to-text transcription or had grading scales which required careful consideration by a human rater. The rating systems also did not explicitly track specific errors. Yang's (2020) rubric simply graded the correct or incorrect pronunciation of individual words allowing explicit conclusions to be drawn.

The present study combined Penkhae's (2020) focus on problematic sounds with Yang's (2020) simplified rubric. The use of Yang's rubric allowed for error diagnosis via speech-to-text transcription and also allowed us to gather data on word and syllable pronunciation errors.

Yang's study also validated the use of the "Rainbow" passage, which was adapted by the Speech Therapy Department of the Rochester Institute of Technology for identifying problematic word pronunciation for people with hearing disabilities. Yang adopted the passage for similar diagnostic purposes for L2 English speakers. Using a set passage in this way is further supported by Saito and Plonsky (2019), who found that measurement errors were significantly lower when learners spoke according to a set script rather than spontaneously. This study likewise used the Rainbow passage (see Appendix A).

Typically, researchers identify problematic areas according to the first language (L1) of the speakers (Dai & Wu, 2021; Inceoglu et al., 2020; Lee, Plonsky and Saito, 2020; Lee,

2021; Penkhae, 2020; Yürük, 2020) and select materials for pronunciation assessment accordingly. This study examined pronunciation errors common to L1 Korean speakers from studies by Paik (1977), Cho (2004), and Bauman (2006) and collated a list (see Appendix B) The Rainbow Passage was selected as the material as it was validated for Korean pronunciation assessment by Yang (2020).

Several studies use sound recordings as the primary method of gathering data (Dai & Wu, 2021; Ha, 2021; Inceoglu et al., 2020; Penkhae, 2020; Yang, 2020). The recordings are then processed using a variety of different methods. Penkhae (2020) transcribed the recordings using the international phonetic alphabet. Dai and Wu (2021) did not transcribe the recordings phonetically but had evaluators listen to and rate the recordings. The present study follows Yang (2020), who used ASR to transcribe the sound recordings phonetically so that the transcription could be analyzed conveniently and quickly without subjective differences between raters. It must be noted that Ma, Henrichsen, Cox, and Tanner (2018) are critical of this approach, as they suggest that suprasegmentals (i.e., stress, rhythm, intonation) are not recorded by ASR but are critical to speaking proficiency.

This study adapted Yang's Rainbow passage rubric and method by narrowing the focus from whole words to commonly mispronounced sounds.

3. METHODOLOGY

3.1. Participants

The participants came from seven classes of a first-year speaking course in practical English at a private university in South Korea. All classes chosen for participation consisted of students of varying English proficiency. Out of a total of 82 participants, 18 were not included in the final data sets due to incomplete data (no second recording was performed). Of the remaining 64 participants, five were removed from the final data sets due to poor sound quality or otherwise damaged recordings. Of the 59 participants used in the final data sets, there was a treatment group consisting of 8 females and 26 males (n=34) and a control group consisting of 13 females and 12 males (n=25). Of this group, 43 completed both the tests (pre-test and post-test) and the post-study questionnaire.

3.2. Procedure

The experiment was conducted in three phases: a pre-test recording, a treatment period, and a post-test recording. Students who completed all three phases of the study were then asked to fill out a questionnaire. The pre-test recording was administered without preparation.

Students were asked to record themselves reading aloud a pre-selected passage (the Rainbow passage) using their smartphone voice recorders. Recordings were uploaded to a shared folder on a cloud storage website.

The original Rainbow passage contained numerous difficult words for Korean speakers to pronounce (i.e., “strikes”, “raindrops”, “above”, “gold”). It was modified slightly by adding the words “changing” and “traveling,” which offered additional opportunities to test problematic sounds. The passage typically took between 40 and 62 seconds to speak aloud. Students were trained in making quality recordings, including advice on speaking at a consistent, loud volume and maintaining separation between mouth and microphone.

All students were asked to practice the Rainbow passage each week during the treatment period. It was explained that after the treatment period students would make a second recording of the passage to be compared to the first one. After four weeks, all students were instructed to make a second recording using their smartphone voice recorders. Students were allowed to make as many attempts as desired, but they were required to complete the entire passage in one attempt (rather than piecing together the “best attempts” from various recordings). The second recordings were uploaded to cloud storage following the same process as for the first recordings.

3.2.1. Pronunciation training

All students received explicit instruction regarding pronunciation differences between English and Korean. When the Rainbow passage was first given to students, their initial approach was to annotate each English word with its equivalent approximation in Hangeul (see Figure 1). The pronunciation instruction began with perceptual training regarding the pronunciation errors arising from this annotation method due to differences between spoken English and Korean. For example, students were shown the difference between ‘strike’ as pronounced in English, which consists of one short syllable, and the transliterated Hangeul equivalent (스트라이크), which expands to five syllables with four consonants, five vowels, and one placeholder. Students were then given training in adapting mouth shapes and tongue positions (see Figure 2) to correct common pronunciation errors. The images shown in Figure 2 were shown to students to illustrate the physical differences between spoken English and Korean. Students were also introduced to common mistakes made by Korean speakers and offered coaching and training in how to avoid these mistakes.

The common error list was divided into parts which were covered during four training sessions, and students were asked to practice with focus on the specific sounds in class for fifteen minutes. They had the opportunity to ask questions and collaborate with classmates in a short pronunciation workshop.

FIGURE 1

Transliteration Examples

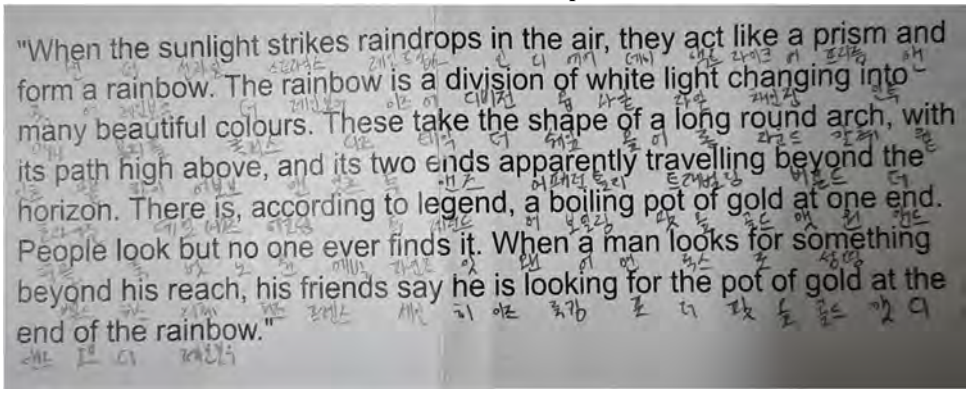
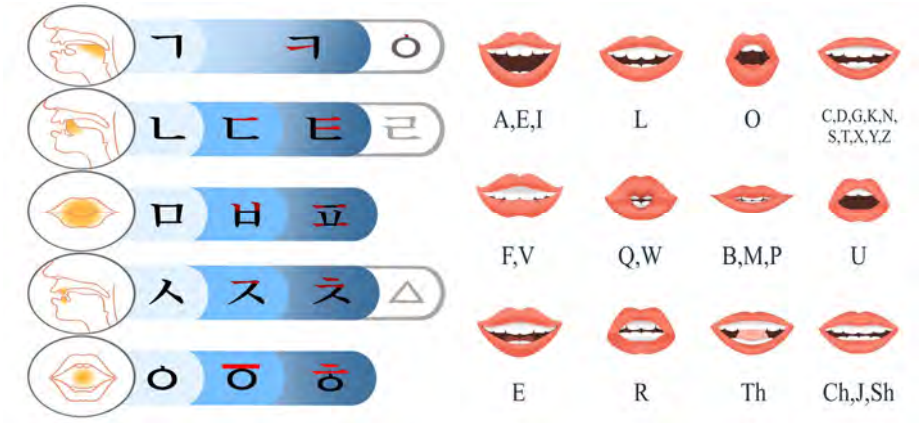


FIGURE 2

Comparison of Korean and English Mouth Shapes



Source: Images used with the permission of Lee Chae Min Graphic Design

3.2.2. ASR training

For the treatment period students were randomly divided into two groups, with one group (the treatment group) receiving detailed training in self-study methods involving the use of automatic speech recognition (ASR) feedback with Google Documents. The other group (the control group) did not receive training in the use of ASR but did go through the same process of pronunciation training described above.

Students in the treatment group were shown how to access ASR using the microphone icon in Google Documents. They were also shown how ASR can be accessed via similar icons present in the mobile keyboard of the Google Translate and Papago applications. Next,

they were given a read-aloud study strategy utilizing ASR. This strategy was as follows: 1) select a sentence from the Rainbow passage for practice, 2) activate ASR and speak the sentence, 3) compare the text output on the smartphone to the selected sentence. If the text output on the smartphone did not match the selected sentence exactly, this presented students with an opportunity to identify specific pronunciation errors they might be making. Students were encouraged to correct these errors and repeat the sentence until there was a perfect match. Students in the treatment group were given 15 minutes to practice during class once a week for four weeks and also encouraged to practice in the same way at home.

To help students use the read-aloud strategy, four important features and limitations of ASR and translation applications were demonstrated. First, students were shown the pronunciation modeling feature present in both the Google Translate and Papago applications. This feature allows students to type in an unfamiliar word and hear the correct pronunciation of that word. Second, it was shown that ASR is typically more effective with full sentences and long phrases as opposed to isolated words taken out of context. In light of this limitation, students were encouraged to use full sentences from the Rainbow passage when practicing. Third, it was shown that ASR can be sensitive to speaking rate and that long pauses can lead the system to assume that the speaker is finished and deactivate. Fourth, the issue of noise and sound interference with recordings was demonstrated and students were encouraged to speak at a moderate volume with their mouths at least several inches away from the microphone. Students were also encouraged to practice in a quiet place with few sound sources around them.

3.3. Questionnaire

After the post-test recording was complete, students were invited to complete an online questionnaire via Google Forms (see Appendix C). The questionnaire collected demographic information (student number, gender, age, and how many years studying English) as well as questions regarding the experiment. Students were asked five questions about studying the passage (5-point Likert scale, Cronbach's alpha of 0.86), and four questions about recording the Rainbow passage. Of these latter four questions, three were 5-point Likert scale questions (Cronbach's alpha of 0.69), and one was a linear scale question "How many times did you record the Rainbow passage?" In addition, the treatment group was asked three questions about using the voice typing (5-point Likert scale, Cronbach's alpha of 0.78). All participants were also asked two linear scale questions regarding how many days they spent practicing and how many minutes they practiced each time, as well as one binary question about studying with a friend or not.

All questions were presented in both English and Korean. The Korean version was trans-

lated from English using Google Translate and then edited, corrected, and verified by a native Korean speaker. A spreadsheet was derived from the Google Forms data which allowed the mean and standard deviation for each of the qualitative items to be calculated.

Demographic information was also collected with the online questionnaire (see Appendix D). Students were asked about how much experience they had studying English ("Years of Experience") as well as how much they studied or practiced each week during the treatment period ("Days studied per week," and "Session time") and whether or not they practiced with a friend ("With a friend"). Finally, students were asked how many attempts they made to record the Rainbow passage the second time ("Number of recordings"). Members of the treatment group and members of the control group were each asked these questions.

Members of the treatment group were also asked questions to assess their engagement with ASR, since for the treatment group "Days studied per week" would include using ASR as instructed in class. Members of the treatment group were asked how often and for how long they studied using ASR. Out of 27 respondents, 15 claimed to study 1 or 2 days per week. Most of the group (n=20) claimed to study between 5 to 10 minutes each time. There was no observable difference in improvement according to days studied per week or session time. Seven members of the treatment group (total n=34) did not fill out the questionnaire. Treatment group participants took some time to record the Rainbow passage correctly the second time, with 24 of 27 reporting that they made at least 3 recordings of the passage on the second pass.

3.4. Data Analysis Methods

All recordings in the pretest and posttest were transcribed using Google Cloud speech to text with the American English accent setting. Transcription confidence values were then checked within the speech to text system. If the confidence rating was low (indicating poor sound quality), the recording volume was normalized using AIMP audio editing software. Five recordings were deemed unusable due to extremely poor or inconsistent sound quality and were eliminated from the data, as mentioned above.

The recordings and transcriptions were then analyzed and errors were categorized according to the pronunciation error list. The full list can be found in Appendix B; however, for the purposes of error checking, several errors were condensed together as follows: contrast between /l/ and /r/, final /l/ and /r/, and /l/ and /r/ within a consonant cluster were combined as LR; contrast between /f/ and /p/, /b/ and /p/, /b/ and /v/, and /f/, /b/, and /v/ were grouped together as BPFV; extra eu & ee and consonant clusters in the initial position were categorized as Epenthesis; all vowel errors were grouped together as Vowel.

In cases where the transcription error was ambiguous or uncertain, the original recording was consulted directly by the researchers to make a determination. If an error could not be

categorized, for example, as clearly a minimal pair error, it was placed in a miscellaneous (MISC) category. For each recording, the total number of segmental errors was calculated to allow for a comparison between pre-test recordings and post-test recordings. The rainbow passage contains 131 syllables; however, some syllables had more than one segmental error. For example, the single syllable word "gold" was commonly mistranscribed as "court," which was counted as three errors: 1 miscellaneous error for the g>c, one vowel error o>ou, and one L/R error for ld>rt). These calculations produced a segmental error rate (SER).

As a hedge against researcher bias, the results generated according to the SER were correlated against scores generated by an online word error rate checker via https://www.amberscript.com/en/wer-tool, which resulted in a reasonably strong Pearson's correlation of .788. All data was then correlated and analyzed with demographic information taken from the questionnaires to look for trends. Data analysis was performed using the Statistical Package for the Social Sciences (SPSS) version 27.

4. RESULTS

This study focused on the effectiveness of using automatic speech recognition (ASR) to improve pronunciation. Participants were divided into a treatment group, which received instruction in and practice time with ASR-based feedback, and a control group, which did not use ASR to practice their pronunciation. All participants made two recordings: one prior to the four-week treatment period and one posterior to it. The results of the two recordings made by each of the groups were then compared to address the research hypotheses.

At the beginning of the study all participants took part in a read-aloud pre-test. All of them took the same read-aloud post-test at the end of the study. For each group, descriptive statistics such as means and standard deviations were calculated. The purpose was to compare the pre-test and post test Segmental Error Rate (SER). Paired samples t-tests were performed to discover whether there was a significant reduction in SER for each group. A one way ANOVA test was administered to discover whether there was a significant difference between the error rates of the two groups.

4.1. Overall Pronunciation Accuracy

The first research question was: "Does guided ASR practice improve Korean L2 English learners' overall pronunciation accuracy?" Our hypothesis was that the treatment group using ASR would make measurably greater gains in overall pronunciation accuracy. This hypothesis was largely confirmed. The test results are presented in Table 1.

TABLE 1

Paired T-test Results on SER Error Rate for Pre and Post Tests

Group	Pre-test		Post-test		*t*	*p*	Cohen's d
	M	*SD*	*M*	*SD*			
Treatment (n=34)	26	11.12	23.03	9.41	3.01	0.01	0.52
Control (n=25)	18.36	6.18	18.4	6.37	-0.53	0.96	-0.01

For the participants in the Treatment group the mean error rate of the pre-test was 26 (*SD* = 11.12) while that of the post-test was 23.03 (*SD* = 9.41). For participants in the control group, the mean error rate was 18.36 (*SD* = 6.18) while it was 18.4 (*SD* = 6.37) for the post-test.

The findings of the paired samples t-tests show that there was a significant reduction in mean error rate with a moderate effect size for the Treatment group ($t = 3.01$, $p = 0.005$, d = 0.52). However, for the control group there was a slight increase in error rate ($t = -0.53$, $p = 0.96$)

Paired t-tests do not assume equality of variance and it was seen that the pre-test means showed a big difference so a Levene's test was conducted to check the homogeneity of variance of pre-test scores (See Table 2)..

TABLE 2

Levene's Test of Equality of Variances on Pre-test Results Between Groups.

	Levene statistic	*df*	*Sig*
Pretest SER	9.215	1(57)	0.004

It was seen that the variance of the pre-test scores of each group did not show a homogeneous distribution (Pretest, $F = 9.595$, $p < .05$). This could be attributed to the unequal sample size of each group and the quasi-experimental design using university sections with mixed abilities.

The error rates for each group were calculated. Descriptive statistics for the SER error rate are reported in Table 3.

TABLE 3

Summary of SER Error Rate

Group	*M*	*SD*
Treatment (n=34)	2.975	5.75
Comparison (n=25)	-0.4	3.8.

The Treatment group ($M = 2.975$, $SD = 5.75$) outperformed the Control group ($M = -0.4$, $SD = 3.8$), however the sample sizes were unequal with heterogeneous variance, so a one-way Welch's test on the SER error rate was conducted in order to determine if the difference

was significant.

TABLE 4

Result of ANOVA Test on SER Error Rate Between Groups

	df	*F*	*p*	*ηp2*
SER error rate between Groups	1(57)	5.176	0.027	0.83
Welch's t-test	1(56.46)	5.845	0.019	-

As shown in Table 4 above there was a statistically significant difference in SER error rate between the treatment and control group with a moderate effect size. ($F(1,57) = 5.176$, $p = 0.027$, $\eta p\ 2 = 0.83$). The Welch adjustment $F(1,56.464) = 5.845$, $p = 0.019$) upheld the significance of the difference between groups despite unequal variance.

4.2. Common Pronunciation Errors

The second research question was: "Does guided ASR practice help Korean L2 English learners avoid common pronunciation errors?" Our hypothesis was that guided ASR practice would help learners avoid these errors.

The results of the treatment group and control group with regard to specific pronunciation errors are shown in Table 5. According to these results, our hypothesis was largely confirmed. Specifically, the treatment group showed improvement with the following errors: LR, J/CH/Z, Epenthesis, S/SH, Vowel (as well as with miscellaneous uncategorized errors). In all cases, these improvements were greater than those which can be found in the control group.

However, there are two limitations with the conclusions that can be drawn from the data. First, some pronunciation errors did not show improvement at all (Wh/F), while others showed regression (BPFV). Second, the results of paired t tests in each category did not demonstrate statistically significant improvement. While improvements did occur, none of the error types listed show a significant *p*-value (which was set at >.05).

TABLE 5

Mean Error Changes in Two Groups

Error Type	Treatment		Control	
	Pre	Post	Pre	Post
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
Vowel	8 (2.24)	7.21 (2.79)	6.52 (2.26)	6.48 (1.76)
LR	4.65 (2.67)	4.12 (2.27)	3.32 (2.29)	3.28 (2.05)
Epenthesis	4.12 (2.82)	3.65 (2.3)	2.96 (1.37)	2.92 (2.31)
MISC Error	3.83 (3.01)	3.39 (1.59)	3.38 (2.5)	3.2 (1.54)
BPFV	3.21 (2.37)	3.26 (2.5)	1.96 (1.81)	1.88 (1.42)

-S/+SH	2.14 (1.2)	2.1 (0.94)	2.07 (1.11)	2.07 (0.92)
J/CH/Z	1.5 (0.82)	1.31 (0.87)	1.43 (0.75)	1.29 (0.78)
Wh/F	1.33 (0.58)	1.33 (0.58)	1.17 (0.41)	1.17 (0.41)

Most common errors were with vowels, L/R and epenthesis which are extremely frequently occurring sounds, while fewer errors were seen with less common sounds Wh/F and J/Ch/Z. Large numbers of errors were noted with monosyllable function words (i.e., "a," frequently mistranscribed as "our," and "or," "the," often seen as "to," "these," as "this" and "its" as "each"). Content words "gold," "pot," and "reach" showed a wider locus of individual errors, (i.e., "gold" pronounced as "court," "god," "good," "pot" as "called" and "part," "reach" as "leech," "rich," and "lich"). Conversely, some of the words that showed great improvement were prism, strikes, legend, sunlight, boiling and path.

4.3. Evaluation of ASR Practice

The third research question was: "How do students evaluate guided ASR practice?" Our hypothesis was that participants would respond positively. This question was addressed by the post-test questionnaire, which posed 11 questions using a 5-point Likert scale (results shown in Table 6). Students responded positively to the pronunciation training, with 90.9% agreeing that the pronunciation training was very good ($M = 4.68$, $SD = 0.639$). The questionnaire also specifically addressed the use of the Rainbow passage and how students responded to the use of it as a testing instrument. The Rainbow passage was felt to be useful, with 84.1% agreeing that the passage was useful, that it helped improve pronunciation, and that it was a good test of their abilities.

More notably, responses were somewhat mixed on the technical issues of conducting the experiment. This was shown in the responses to such questions as "It was easy to study the Rainbow passage," ($M = 3.52$, $SD = 0.762$), "Voice typing was reliable on my smartphone" ($M = 3.63$, $SD = 0.967$), "Recording the Rainbow passage was easy" ($M = 3.64$, $SD = 0.917$), and "Recording the Rainbow passage was comfortable" ($M = 3.7$, $SD = 0.817$). On the question of whether "Recording the Rainbow passage was better than a face-to-face test," students generally agreed ($M = 3.75$), but with an extremely large variance in responses ($SD = 1.164$).

TABLE 6

Response to Questionnaire Likert Scale Items

	Mean	*SD*	Net Agree %
It was easy to study the rainbow passage	3.52	.762	50
The training for pronunciation was good	4.68	.639	90.9
The Rainbow passage was a useful way to study	4.19	.664	84.1
The Rainbow passage helped me improve	4.20	.701	84.1

The Rainbow passage was a good test	4.18	.691	84.1
Voice typing was reliable on my smartphone	3.63	.967	51.9
I will use voice typing to practice in the future	3.74	.526	70.3
I feel happy about studying using my smartphone	3.93	.730	70.3
Recording the Rainbow passage was easy	3.64	.917	61.3
Recording the Rainbow passage was comfortable	3.73	.817	54.5
Recording the Rainbow passage was better than a face-to-face test	3.75	1.164	61.3

5. DISCUSSION

This study compared the improvement in pronunciation accuracy of two groups of Korean university students. One group was given four weeks of self-study time using ASR technology for evaluation and feedback while the other group was providing a control. In keeping with prior studies (i.e., Inceoglu et al., 2020; Spring & Tabuchi, 2022), this study found modest improvement in pronunciation accuracy with the treatment group as compared to the control group. Although there was no statistically significant improvement with any particular type of pronunciation error, the technology as used in this experiment was able to identify the particular errors which each individual user was prone to make. This suggests that ASR may indeed have benefit as a diagnostic tool for teachers who do not have time to interview and analyze the speech of individual students.

Although the treatment group made improvement in pronunciation accuracy, there was only a moderate effect. This may have been due to the lack of actual practice time with the ASR. As noted above, members of the treatment group studied on average about twice a week for 5 to 10 minutes each time, which may not be sufficient time with the technology to produce much benefit. Also, the treatment period for this study was four weeks, whereas a longer period (i.e., a full semester as per Spring & Tabuchi, 2022) may produce a lower error rate.

ASR training by the treatment group led to a reduction in error rates. This can be attributed to the perception that smartphone use is socially acceptable, and that ASR feedback is objective and private. This study combined comfortable smartphone use with explicit instruction that focused on difficulties common to all Korean learners which were easy to explain and understand with reference to Hangul and possibly further leveraging feelings of social inclusion. We saw that nearly half of the treatment group (N=13) of ASR users practiced with a friend, which may have had extra benefit. This is in keeping with Dai and Wu's (2021) study on peer feedback.

While the Word Error Rate tool was not able to provide detailed error diagnosis, it did correlate well with the SER and can be considered to produce a reliable measure of pronunciation accuracy. This suggests the WER tool could be useful in allowing an error score to be generated quickly and conveniently, even by students. Since the WER is a computer-

generated assessment of pronunciation accuracy with an algorithm that outputs a score for each transcript by tracking variations (substitutions, deletions, and insertions) from the original text, it is easy to generate and may provide a quick heuristic for students to use in self-assessing their progress while providing an efficient data gathering method for a multi stage repeated measures analysis.

Another result concerned the most common pronunciation errors identified by the test. Yang (2020), who also used the Rainbow passage to study pronunciation errors, suggested that errors made with function words were the most likely to be overlooked by human evaluators. Our study found that errors with function words (i.e., "a," "the," "these," and "its") were the most commonly mispronounced. This suggests that pronunciation training could well focus on short one-syllable words that are commonly misspoken in order to provide increased benefit to speakers.

ASR was able to identify errors made with content words as well. The most commonly mistranscribed content words in our dataset were "gold" (172 errors) and "part" (160 errors). These words both appear twice in the passage, making errors more likely to be discovered with them. These results further support the utility of using ASR as a diagnostic tool. They also suggest that, although the G/K error was not included in our original list of common errors it perhaps should have been.

Technical difficulties and limitations may have played a role in the limited results found by this study. While most participants agreed that they enjoyed the pronunciation training and felt that studying the Rainbow passage was helpful (as per Table 7), and that they would use it again ("I will use voice typing to practice in the future"), the more mixed reviews of the technical aspects of the experiment are worth bearing in mind. The researchers believe the Rainbow passage and its sentences may have been too long and difficult to record in one attempt. This difficulty and length may be the reason for the lower ratings from participants. Recording sentences one by one may be a possible solution.

6. CONCLUSION

Automatic speech recognition (ASR) technology offers teachers and students a variety of opportunities to improve language learning. Given the relative lack of opportunity for guided pronunciation practice in many ESL and EFL classrooms, the possibilities held out by ASR for offloading some of this work by technological means seems attractive. If ASR can be shown to have practical benefits with improving pronunciation, its use by both students and teachers is likely to increase significantly in the future.

In keeping with previous studies of ASR effectiveness, this study found that guided practice with the technology did produce a small but measurable improvement in pronunciation

accuracy. By comparing the results of a treatment group which made use of ASR to practice against those of a control group which did not use ASR, we found that the group which did use the technology made gains which the control group which did not. Although specific pronunciation errors were not addressed by ASR, its use seemed to provide a holistic benefit. Furthermore, participants were largely positive about their experience and their intention to make use of it in the future.

This study had several limitations. First, because of the use of students from assigned courses the study cannot claim to be a true random sample and is therefore of a quasi-experimental design. Second, due to the moderately low sample size (n=59), the statistical results were subject to outlier issues—a fact which can be seen in some of the large standard deviation results which were found. Finally, the treatment period of four weeks may have been too small a period for the real benefits of ASR use to become clear. As some participants only practiced once or twice a week over a four-week period, this treatment time may not be sufficient.

Future research should address some of the possibilities suggested by existing studies on the use of ASR technology. For example, what is an optimal treatment time after which measurable benefits could be identified? How much practice with ASR should learners make on a weekly basis to maximize its benefits? How can suprasegmental features be included in analysis of ASR recordings and transcriptions? Also, can ASR be used by learners directly as a diagnostic tool for self-improvement?

Applicable level: Tertiary

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*(4), 778-786. https://doi.org/10.1111/bjet.12354

Ashwell, T., & Elam, J. R. (2017). How accurately can the google web speech API recognize and transcribe Japanese L2 English learners' oral production? *JALT CALL Journal*, *13*(1), 59-76.

Baker, A., & Burri, M. (2016). Feedback on second language pronunciation: A case study of EAP teachers' beliefs and practices. *Australian Journal of Teacher Education*, *41*(6), 1-19.

Bauman, N. R. (2006). A catalogue of errors made by Korean learners of English. In J. Kimball & D. E. Shaffer (Eds.), *KOTESOL Proceedings 2006* (pp. 167-176). Seoul: KOTESOL. Retrieved December 31, 2022 from http://koreatesol. org/sites/default/files/pdf_publications/KOTESOL-Proceeds2006web.pdf

Chen, Z., & Goh, C. (2011). Teaching oral English in higher education: Challenges to EFL teachers. *Teaching in Higher Education*, *16*(3), 333-345.

Cho, B. E. (2004). Issues concerning Korean learners of English: English education in Korea and some common difficulties of Korean students. *The East Asian Learner, 1*(2), 31-36.

Dai, Y., & Wu, Z. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.1952272

Daniels, P., & Iwago, K. (2017). The suitability of cloud-based speech recognition engines for language learning. *JALT CALL Journal*, *13*(3), 229-239.

Dillon, T., & Wells, D. (2021). Student perceptions of mobile automated speech recognition for pronunciation study and testing. *English Teaching*, *76*(4), 101-122.

Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, *35*(8), 1869-1889.

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York: Harper & Row.

Ghoorchaei, B., & Rahmani, M. (2018). The effect of using GET on Iranian intermediate EFL learners' pronunciation and motivation. *Asian Journal of University Education*, *14*(2), 1-17.

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*(1), 70-105.

Ha, J. (2021). Individual differences in perception toward English pronunciation development with ASR technology. *Multimedia-Assisted Language Learning, 24*(3), 10-34.

Hayati, A. M. (2010). Notes on teaching English pronunciation to EFL learners: A case of Iranian high school students. *English Language Teaching*, *3*(4), 121-126.

Holliday, A. (2006). Native-speakerism. *ELT Journal, 60*(4), 385-387.

Hsu, L. (2016). An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning*, *29*(5), 881-900.

Inceoglu, S., Lim, H., & Chen, W. H. (2020). ASR for EFL pronunciation practice: Segmental development and learners' beliefs. *Journal of Asia TEFL*, *17*(3), 824-840.

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception-vs. production-based pronunciation instruction. *System*, *88*, 102185

Lee, Y. (2021). Effectiveness of mobile assisted pronunciation training in the acquisition of English vowels, /o/ and /ɔ/. *Foreign Languages Education, 28*(3), 53-76.

Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge, England: Cambridge University Press.

Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics* (pp. 245-270). London: Palgrave Macmillan.

Lindemann, S., & Subtirelu, N. (2013). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning*, *63*(3), 567-594.

Liu, X., Xu, M., Li, M., Han, M., Chen, Z., Mo, Y., & Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. *International Journal of Innovation and Learning*, *25*(2), 126-140.

Ma, R., Henrichsen, L. E., Cox, T. L., & Tanner, M. W. (2018). Pronunciation's role in English speaking-proficiency ratings. *Journal of Second Language Pronunciation*, *4*(1), 73-102.

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, *5*(1), 98-118.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 617-637.

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.

Paik, Y. J. (1977). Pronunciation difficulties of Korean students learning English as a second language. *Eo Hag Yeon Gu, 13*(2), 177-189.

Penkhae, W. (2020). Improving the Thai students' ability in English pronunciation through mobile application. *Educational Research and Reviews*, *15*(4), 175-185.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652-708.

Spring, R., & Tabuchi, R. (2022). The role of ASR training in EFL pronunciation improvement: An in-depth look at the impact of treatment length and guided practice on specific pronunciation points. *Computer Assisted Language Learning*, *23*(3), 163-185.

Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: Pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, *11*(3), 74-91.

Yang, B. (2020). An evaluation of Korean students' pronunciation of an English passage by a speech recognition application and two human raters. *Phonetics and Speech Sciences*, *12*(4), 19-25.

Yulia, M. F. (2018). Using performance assessment with EFL learners in pronunciation class. *The Asian EFL Journal*, *20*(1), 47-56.

Yürük, N. (2020). Using Kahoot as a skill improvement technique in pronunciation. *Journal of Language and Linguistic Studies, 16*(1), 137-153.

APPENDIX A

The Rainbow Passage

When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light changing into many beautiful colours. These take the shape of a long round arch, with its path high above, and its two ends apparently travelling beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

(Adapted from Fairbanks, 1960. Voice and articulation drillbook, 2nd ed. New York: Harper & Row. pp. 124-139)

APPENDIX B

Common Errors

Contrast between /l/ and /r/
Contrast between /wh/ (hw) and /f/
Contrast between /f/ and /p/
Contrast between /b/ and /p/
Contrast between /b/ and /v/
Contrast between /z/ and /s/ or /j/ (J vs ch~z)
Contrast of /f/ /v/ /b/
Contrast between /l/ and /r/ as final sound
Contrast between /1/ and /r/ as in consonant cluster
Dropping s or s>sh
Extra eu & ee
Minimal Pair /i:/ and /ɪ/
Minimal Pair /u:/ and /ʊ/
Minimal Pair /a/ and /ɔ/ +
/Ↄ/ and /o/ (Long o **/ɔʊ/ or /oʊ/** ~short o **/ɑ/)**
Short a **/æ/** ~ e **/e/ /ɛ/**
Consonant Clusters initial position (fr. skr. by. pl)

APPENDIX C

Questionnaire Items

Domain Item	Item N.	Question
Demographics	Q1.	What is your student number?
	Q2.	Are you male or female?
	Q3.	How old are you?
	Q4.	How many years have you studied English?
Passage	Q5.	It was easy to study the rainbow passage
	Q6.	The training for pronunciation was good
	Q7.	The Rainbow passage was a useful way to study
	Q8.	The Rainbow passage helped me improve
	Q9.	The Rainbow passage was a good test
ASR	Q10.	Voice typing was reliable on my smartphone
	Q11.	I will use voice typing to practice in the future
	Q12.	I feel happy about studying using my smartphone
Practicing	Q13.	How many days did you practice in a week?
	Q14.	How many minutes did you practice each time?
	Q15.	It helped to practice with a friend
Recording	Q16.	Recording the Rainbow passage was easy
	Q17.	Recording the Rainbow passage was comfortable
	Q18.	Recording the Rainbow passage was better than a face-to-face test
	Q19.	How many times did you record the rainbow passage to get it right?

APPENDIX D

Questionnaire Demographics

	Treatment		Control	
	N(%)	Mean(SD)	N(%)	Mean(SD)
Gender				
F	8 (23.5)	4.62 (6.12)	13(52)	0.85(4.56)
M	26 (76.5)	2.46 (5.66)	12(48)	-1(2.63)
Years of Experience				
N/A	7(20.6)	1.86(7.34)	9(36)	-2.56(2.07)
1	3(8.8)	3.33(3.22)	1(4)	1(-)
2	5(14.7)	4.8(7.01)	0	-(-)
3	4(11.8)	2(1.83)	5(20)	1.6(5.81)
4	15(44.1)	3.07(6.04)	10(40)	1.3(3.13)
Days studied per week				
N/A	7(20.6)	1.86(7.34)	2(8)	2.5(7.78)
1	5(14.7)	5(5.05)	1(4)	0(-)
2	10(29.4)	3.3(5.74)	5(20)	0.4(3.13)
3	7(20.6)	3.14(6.44)	4(16)	4.5(3)
4	5(14.7)	1.6(4.51)	2(8)	3(1.41)
5	0(0)	-(-)	2(8)	-2.5(0.71)
Session Time (minutes)				
N/A	7(20.6)	1.86(7.34)	9	-2.56(2.07)
No Practice	0(0)	-(-)	3(12)	0(7)
~5	11(32.4)	3.82(6.6)	5(20)	2.6(2.88)
~10	9(26.5)	2.11(5.01)	5(20)	0.6(4.28)
~15	5(14.7)	3.6(4.93)	1(4)	1(-)
over 15	2(5.9)	4.5(2.12)	2(8)	2.5(0.71)
Practiced with a Friend				
N/A	7(20.6)	1.86(7.34)	9(36)	-2.56(2.07)
No	14(41.2)	3.34(5.85)	12(48)	1(4.39)
Yes	13(38.2)	3.64(5.06)	4(16)	2.5(1.29)
Number of recordings				
N/A	0(0)	-(-)	9(36)	-2.56(2.07)
2	3(8.8)	1.5(2.12)	2(8)	-1.5(3.61)
3	8(23.5)	3.25(5.83)	3(12)	3(5)
4	7(20.6)	2.57(4.76)	5(20)	1.6(4.45)
5	2(5.9)	2(1.41)	2(8)	-0.5(2.12)
Over 5	7(20.6)	6.14(5.84)	4(16)	2.25(4.11)