Student Perceptions of Mobile Automated Speech Recognition for Pronunciation Study and Testing

Thomas Dillon and Donald Wells[*]

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

This study utilized Automated Speech Recognition technology to determine the potential utility and acceptance of such technology in the English as a Foreign Language classroom. Learners were made aware of the Automatic Speech Recognition potential of their mobile devices and provided with some direction in, and incentive for, its use. Participants were then scored on their assessment of the technology according to the Technology Acceptance Model. Participants showed a marked appreciation for the ease and utility of the technology with over 72% agreeing that the technology was both accessible and useful. Support for the use of Automatic Speech Recognition as a testing method was somewhat mixed, with 75% of participants agreeing that the testing was fair, but only 60% reporting that they felt they did well on the test. As a secondary point of interest, this study examined the potential use of Automatic Speech Recognition technology for teaching and testing pronunciation.

Key words: automated speech recognition, ASR, technology acceptance model, TAM, mobile assisted language learning, MALL, pronunciation, EFL

1. INTRODUCTION

Pronunciation is an important part of intelligible language use; however, pronunciation accuracy is an extremely common problem amongst language learners, and discrimination against speakers without native-like pronunciation is a recognized problem among researchers. Pronunciation has also been relatively neglected among researchers of language acquisition. Many reasons have been proposed for this neglect: lack of time, lack of training, deficiency in confidence, and limited resources. Conventional techniques for providing pronunciation instruction have been developed, but such techniques may now be supplemented through technological means.

Computer-assisted language learning offers an increasingly popular and potentially useful mode of improving pronunciation. Some studies have been critical of the accuracy of computers versus human listeners, but as speech recognition technology has matured—and as the availability of speech recognition technology has rapidly expanded, due in part to the proliferation of smartphones—a closer look at the use of speech recognition technology is in order (Hsu, 2016).

Research has shown that some Automated Speech Recognition (ASR) learning methods can be effective for pronunciation practice (McCrocklin, 2019a). ASR practice has many noted benefits: it is adaptable to individual learning styles; it is free-form (meaning topic and scope of learning are not limited); it is convenient, cheap, mobile, and accessible; and it is private. However, many language learners are unaware of the opportunities to use ASR provided by most ordinary smartphones. Many are unaware of the various applications available to them which have ASR use built in as a standard.

This study looked at the responses from English language learners at a Korean university once they had been made aware of the possibility of ASR use for pronunciation practice and given instruction in its use. To encourage the use of ASR, participants were also given a pronunciation test for which the ASR practice could potentially improve the results. This study had three research questions:

1. What responses do learners have to using ASR for pronunciation practice?
2. Which type of application would participants consider the most useful?
3. How do learners feel about using ASR as a testing method?

2. LITERATURE REVIEW

2.1. Pronunciation in Language Learning

Modern approaches to language teaching aim at achieving an intelligible accent rather than creating native-like pronunciation (Levis, 2005). This move has been prompted in part by critiques of "native speakerism" (an example of such a critique appears in Holliday, 2006). While this change of focus has lowered the bar for pronunciation skill acquisition, many listeners still have great difficulty in understanding accented speech (Harding, 2011). Furthermore, achieving an intelligible accent remains a difficult goal, particularly for learners without access to immersion in a second language (L2) environment or significant amounts of L2 interaction (Baker & Burri, 2016).

McCrocklin (2016) has shown that pronunciation instruction can have positive effects in low-L2 exposure settings, but there are many practical obstacles to the application of this observation. For example, pronunciation improvement is often sidelined in favor of what is perceived by instructors as a more efficient use of class time (Couper, 2017). Some classroom situations can even be perceived as overwhelmingly difficult for giving pronunciation instruction, with the example of a 50-student Chinese classroom being discussed in detail by Liu at el. (2019). Another obstacle comes from teachers themselves who can find that explicit pronunciation correction may be hurtful to student feelings (Baker & Burri, 2016; Couper, 2017).

2.2. Using Technology to Teach Pronunciation

Technological assistance seems to offer a way to provide the pronunciation instruction that is often lacking in the classroom. Computer Assisted Pronunciation Training (CAPT) was one of the first systems to attempt this, and its utility in providing highly specific feedback for particular pronunciation errors is widely recognized (McCrocklin, 2019a; Mroz, 2018; Neri, Cucchiarini, & Strik, 2008; Wang & Young, 2015). The drawbacks with CAPT lie in its difficulty of use (Daniels & Iwago, 2017; Wang & Young, 2015), lack of adaptability (McCrocklin, Humaidan, & Edalatishams, 2019), and cost (McCrocklin, 2018).

More recently, ASR has emerged as a technological option for providing pronunciation assistance (Hsu, 2016). Advancements in "Artificial Intelligence" (AI) natural language processing have improved ASR and allow it to output results as text which even non-expert users can understand (Daniels & Iwago, 2017). Studies have shown that ASR can be equal in effectiveness to classroom instruction when it comes to pronunciation practice (McCrocklin, 2019a). ASR has also proven useful in situations when time with first language (L1) users is limited (Kim, 2006). While there were issues with the accuracy and validity of

ASR systems for judging native and non-native speech, those concerns have disappeared as the technology has matured, and the modern opinion is that ASR judges speech about as well as a human being (Evanini, Hauck, & Hakuta, 2017). One detailed analysis of the transcription ability of Google voice typing returned results showing a recognition of native speech at a level almost equal to a human, with a recognition of non-native speech about 3-5% lower (McCrocklin & Edalatishams, 2020). Even if granted that ASR is imperfect, Guskaroska (2019) has shown that learners can still benefit even from imperfect feedback.

Using ASR offers many benefits to learners, particularly given the increasing availability of mobile technologies. These benefits include speed, real-time transcription, unbiased feedback which does not interrupt speech, and ease of use (Daniels & Iwago, 2017; McCrocklin, 2018; Mroz, 2018). Also, because ASR can be less sensitive to accent, learners can achieve comprehensibility without the need to conform to native-like pronunciation (Evers & Chen, 2021). In one study (Mroz, 2020) ASR users significantly improved their intelligibility and proficiency as compared to non-ASR users. Modern ASR systems allow users to practice with any topic or vocabulary, showing a level of flexibility far beyond earlier systems such as CAPT (Evers & Chen, 2021). Useful ASR functions include speech to text transcription, audio input to provide feedforward, the ability to look up words and get lists of similar words, and tips on producing sounds. (Liakin, Cardoso & Liakina, 2017; McCrocklin, 2018). The current study makes use of ASR dictation software to provide feedback to learners. This use of ASR has been studied by McCrocklin (2019a, 2019b) and Shadiev, Hwang, and Huang (2014).

Studies also raise the possibility of using ASR systems in formative assessments. The current study is a step toward this. Previous studies in this vein include McCrocklin (2018) and Zechner at al. (2015). Fully automated L2 speaking tests have been available since 2015 with Pearson's Versant English Test (Isaacs, 2017). A test of English as a foreign language (TOEFL) preparation system, "Speechrater," has been considered useful by both teachers and learners (Gu, Davis, Tao, & Zechner, 2021). The computer-based English Learning and Speaking Test (CELST) uses "Iflytek" scoring technology (Liu et al., 2019).

Using ASR is not without its problems. Some applications may lack the necessary functions for phonetic description, making it difficult for users to understand how to vocalize sounds properly (Evers & Chen, 2021). ASR dictation does not improve listening skills (Liakin, Cardoso & Liakina, 2015). Some users remain unable to differentiate between their pronunciation and native pronunciation, making changes more difficult (Garcia, Kolat, & Morgan, 2018). Teacher support may still be necessary for providing strategies and encouragement (Moyer, 2014). This last finding is particularly relevant, as it suggests a continuing role for the teacher, even in an environment of technology assisted learning (Evers & Chen, 2021). Despite such problems, researchers have promoted the use of ASR on multiple grounds. One example is McCrocklin (2019a), who argues that the flexibility of

the technology outweighs feedback limitations and the possibility of errors. Neri et al. (2008), discussing CAPT systems, found that simply indicating mispronunciations was a benefit to learners even without other detailed feedback.

Furthermore, existing research on Automated Speech Recognition suffers from several notable limitations. One limitation is the participant pool, which is often small and composed entirely (or almost entirely) of highly motivated learners. Mroz's work is a case in point, as both the 2018 and the 2020 studies involve small groups (16 and 26, respectively) of highly motivated learners in a foreign language degree program aimed at achieving advanced proficiency. Another example is the 2017 study by Liakin, Cardoso and Liakana, which included 69 participants (of which only 14 participated in the Automated Speech Recognition exercises) who were motivated learners in an advanced degree program. Our study aims to address this by looking at a large pool of participants representing various levels of English language proficiency.

Another limitation concerns the use of translation applications with Automated Speech Recognition capabilities. Research has only recently started to acknowledge the usefulness of translation applications for language study (Amin, 2020; Sagita, Jamaliah, & Balqis., 2021; Yadav, 2021). One advantage noted by Mroz (2020) concerns the use of translation programs which have been trained in the use of different accents. Mroz observes that this allows learners to focus on developing intelligibility rather than conforming to a standard accent. In order to refocus on the learner, participants in our study were given free choice of which Automated Speech Recognition application to use. The majority of participants showed a preference for translation applications (specifically Naver and Google).

One additional limitation of existing research which bears mentioning is the relative under-reporting of the use of Automated Speech Response in testing or the attitudes of learners toward such use. While our study does not examine the pedagogical effectiveness of Automated Speech Response in testing, it does provide some data on how users respond to its use. Teachers interested in the use of Automated Speech Recognition for testing may be interested to see further studies along these lines which could further parse the responses of learners to Automated Speech Recognition-based testing.

2.3. User Responses to Technology Assisted Learning

The use of ASR is somewhat different when considered from the perspective of learners. Several studies have looked at individual practice with ASR. The responses are often positive. In particular, users have tended to find using ASR for pronunciation practice useful, enjoyable, and uncomplicated (Cucchiarini, Neri, & Strik, 2009; Guskaroska, 2019; Liakin et al., 2017). Users have also been largely satisfied with the ease of use of ASR (Evers & Chen, 2021). Negative attitudes toward ASR have largely centered on frustrations due to

incorrect feedback (Kim, 2006). Among the errors leading to user frustrations, several stand out: splitting long words incorrectly into several short words, system shutdown during user pauses, system failing to recognize context, and feedback output vocabulary that was unknown to the user (Liu et al., 2019). Such problems are associated with decreased motivation among users (Liakin et al., 2017; McCrocklin, 2018).

In general, technology use such as Mobile Assisted Language Learning (MALL) has increased among learners to such a degree as to put pressure on instructors. Mobile devices such as smartphones have become so ingrained in the lives of users that pedagogical approaches must explore the opportunities provided by machine translation or ASR (Ducar & Schocket, 2018). Recent studies into mobile device ownership and technology acceptance show that learners accept teachers to support students' adoption of mobile learning resources (Hoi & Mu, 2021). Web 2.0 applications such as social media and translation applications such as Google Translate also have become widely expected and accepted by language learners (Yadav, 2021). There is even evidence that the use of such applications can lead to improvement in language acquisition (Amin, 2020; Sagita et al., 2021).

Given that language learners now expect to be able to use applications and mobile devices in a classroom setting, there seems to be an opportunity for teachers to provide encouragement, instruction, and direction. With the acceleration in interest in distance learning and mobile learning during the pandemic, there is a pressing need to better understand the perceptions and acceptance of new technologies.

3. METHODOLOGY

3.1. Participants

The participants came from five class sections of a first-year Practical English course at a private university in South Korea. All classes chosen for participation were composed of students of varying English proficiency studying a variety of majors. The classes were selected to examine the responses of students from a diverse range of English proficiency. A total of 211 students participated in the study and completed the questionnaire. Participants were first-year students composed of 146 males (69.2%) and 65 females (30.8%) studying 27 different majors ranging from Engineering (28.4%) and Visual Design (12.8%) to Software, (7.6%) Education (7.1%) and Languages (6.2%).

3.2. Procedure

The study consisted of a three-stage program and a post-program questionnaire. The three stages of the program were: 1) Introduction and Demonstration, 2) Self Study Period, and 3) Low Stakes Mock test. The details of each stage are given below.

3.2.1. Stage 1: Introduction and demonstration

The first stage of the program took place at the beginning of the semester. Participants were introduced to the Automated Speech Recognition (ASR) capabilities of their smartphones and shown how to apply this technology to produce feedback highlighting pronunciation mistakes. ASR functionality is standard on most modern smartphones, and each participant involved in the study owned or had access to such a smartphone. While there are tailored language-learning ASR applications available, many of these were developed for research or commercial use and require a paid subscription (McCrocklin, 2019a). To encourage greater participation, therefore, participants were given the chance to use free applications of their choice. One of the questionnaire items asked participants to identify the application(s) they chose to use. This initial session lasted for around 15 minutes and focused on enabling ASR on each participant's smartphone and then encouraging participants to try out the technology freely.

ASR is specifically accessible through any application which uses a mobile keyboard. Typically, the mobile keyboard will have a toolbar from which one may select ASR to replace text entry (see Figure 1). Some individual instruction took place to allow participants with different operating systems to install or enable ASR on their smartphones.

FIGURE 1
ASR Icons Android (Top) Apple (Bottom)

Source: Screen capture

The following week after ASR had been enabled, another 15 minute session was conducted during which a simple read-aloud study strategy was demonstrated. This strategy was as follows: 1) select a sentence from the course textbook for practice, 2) activate ASR and speak the sentence, 3) compare text output on the smartphone to the selected sentence. If the text output on the smartphone did not match the selected sentence exactly, the sentence was repeated until there was a perfect match.

Through the application of this strategy learners became aware of potential pronunciation errors and learned to identify specific areas of improvement. In the sample selection, short words such as 'in', 'on', and 'an' were a common source of pronunciation errors. Having the errors displayed in text form guided efforts toward improved pronunciation.

To further aid in the successful adoption of the read-aloud strategy, participants were made aware of a useful smartphone feature: the ability to type a word and hear it pronounced clearly. It was observed that some participants were uncertain of the correct pronunciation of certain words in their textbook, leading to a persistent repetition of pronunciation errors. To obviate this issue, participants were encouraged to use their smartphone dictionary or translator to model the correct pronunciation of unfamiliar words.

As part of the demonstration, participants were also made aware of certain limitations of this approach. First, it was demonstrated 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, participants were encouraged to repeat an entire phrase or sentence containing an error rather than simply to repeat a single word which had been mispronounced. Next, ASR can be sensitive to speaking rate such that long pauses can lead the system to assume speaking is completed. Participants were encouraged to overcome speaking hesitancy to counter this issue. Finally, classroom participants were observed to whisper closely into their smartphone microphones. It became necessary to explain that this method could overpower the microphone, producing noise which interfered with the ASR capabilities. To overcome this issue participants were encouraged to speak at a moderate volume with their mouths at least several inches away from the microphone.

Participants were given only minimal guidance after the installation and demonstration to allow more freedom to develop personalized study strategies for use in Stage 2.

3.2.2. Stage 2: Self-study

The second stage of the program was a 14-week period of self-study. For this stage, participants were invited to develop their own ASR pronunciation strategies and apply them as they saw fit. Participants were informed that at the end of the self-study period they would take a pronunciation test. They were also informed that the test itself would not be graded and had no bearing on their course evaluation.

Two weeks before the scheduled test, there was a third 15 minute session where participants were encouraged to practice sentences from the course textbook and given a document containing practice material. The practice material consisted of 60 sentences taken from the course textbook (material which was covered in class during the ordinary run of class time). Participants were made aware that the Low Stakes Mock Test at the end of the self-study period would be a review of the practice material.

3.2.3. Stage 3: Low stakes mock test

The third and final stage of the program was a simulated test of the studied material. This test served multiple purposes: it provided a challenge to allow participants to analyze the utility of their ASR practice method, and it allowed the researcher the opportunity to infer which participants had taken the time to practice self-study with ASR.

The test conditions were as follows: participants were given a list of 30 sentences selected from the pool of 60 study sentences assigned during Stage 2. The sentences were displayed as a numbered list on a single sheet of paper. From this list a set of five sentences was randomly assigned to each participant at the beginning of the testing period. The goal of the test was for each participant to pronounce each of their five assigned sentences correctly while under observation by the researcher. To help with possible test anxiety, participants were reminded beforehand that the test would have no impact on their course score. These guidelines were explained to each participant at the beginning of the testing period. All participants used the same computer and microphone for data collection.

The testing procedure was as follows: each participant was given instructions on how to complete the test and assigned five questions from the list of 30 by random number generation. Microphone placement was arranged to facilitate ease of recording and the researcher monitored each participant closely to ensure that none got too close to the microphone. Participants were then shown five sentences on a computer screen with an icon displaying when ASR was active in recording their speech. Participants were instructed to pause after each sentence to allow the researcher to collect the results in a Google document.

For participants whose speaking volume was too low, or whose speed was slow and hesitant, the researcher interrupted the test to provide corrective feedback and ensure that technical problems did not mar the overall results. In some cases where the ASR system did not display the complete sentence due to technical or audio difficulties, the researcher allowed participants to repeat some or all of their sentences. Participants were allowed to re-record sentences in cases where they self-diagnosed a pronunciation error. Participants were also allowed to read through their answers before completing the test and re-record sentences which they perceived as improperly pronounced.

The testing procedure was not uniform in that participants were permitted to make

multiple attempts at successfully pronouncing the sentences. Some participants discontinued after their first attempt, while others made multiple attempts in an attempt to improve performance. This procedure was adopted to encourage compliance and reduce testing anxiety. Due to the non-uniform nature of the procedure, however, data from the pronunciation test results was saved as a Google document but not analyzed. Some observations concerning this procedure may be found in the Discussion section below.

3.3. Instrument

Following the three-stage program above, participants were given a 10-item questionnaire. The questionnaire was designed to gauge student response to the program and test and to elicit feedback. The questionnaire included eight quantitative items, with three concerning student use of ASR during the self-study period, four concerning the application of ASR to the test in the testing period, and one concerning the teacher's perceived fairness in applying the system to all participants. These questions were adapted from Hsu's (2016) Technology Acceptance Model questionnaire, with changes to ensure comprehension from participants of varying English proficiency. These items used a 5-point Likert scale and had a Cronbach's alpha of 0.86, suggesting relatively high consistency. On this scale, a 1 represented high disagreement with the statement and a 5 represented high agreement. The questionnaire included two qualitative items. One item asked which application or applications the participant chose to use for self-study. The other item was for open-ended responses for "My comments." A total of 104 participants left comments.

The eight quantitative items on the questionnaire examined three main measurement constructs. The three items concerning student use of ASR during the self-study period examined Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), which are the two main dimensions of the TAM as used by Weng, Yang, Ho, and Su (2018). This set of questions showed good internal consistency (Cronbach's alpha 0.82). Four of the remaining five items examined Attitude, (AT) via sub-constructs Technology Anxiety and Self Efficacy, which are features of the extended Technology Acceptance Model of Zheng and Li (2020). As AT is considered to be a synthesis of PU and PEOU, questions measuring AT do not need to directly reference Ease or Usefulness. We chose to measure AT in order to produce complex but easily intelligible questions related to the testing procedure. Weng, Yang, Ho, and Su (2018) similarly adapted the Technology Acceptance Model to reference other concepts and ask for comparisons. The last remaining item was designed to examine the sub-construct of Family Support. This item asked about the effect of teacher support during the test, as the "Human in the Loop" is an important support feature of modern AI systems (Zanzotto, 2019). As per Zheng, and Li (2020), we extended the concept of "family support" to include the idea of teacher support for this item. The five items regarding the test

had good internal consistency (Cronbach's alpha 0.77). The quantitative results were further corroborated by qualitative comments. Thematic coding of the qualitative results showed 83.6% positive AT, with the largest percentage of positive AT-related comments (27.9%) directly referencing the testing method.

The raw data from the paper questionnaires was carefully input into Google Forms. A spreadsheet was generated for each of the eight quantitative items to find the mean scores and the standard deviation for each set of responses. The full results of the data may be found in Appendix B in Table 7. The open-ended responses were also input into Google Forms and analyzed using the constant comparative method. Results were categorized inductively into "general comments" (comments that could not be attributed to a specific aspect of the process) and "specific comments" (comments that directly referenced some aspect of the process - for example, the test). All comments were then classified according to the Technology Acceptance Model (TAM) and are included in the Discussion section below. The items are listed in Appendix A in Table 6.

4. RESULTS AND DISCUSSION

4.1. Results

This study focused on user responses to using ASR for pronunciation study. Questions regarding the pedagogical value of ASR were not specifically addressed, although how users perceived that value was. The sections below will review the research hypotheses and present the corresponding results for each.

4.1.1. What responses do learners have to using ASR for pronunciation practice?

The first research question was: "What responses do learners have to using ASR for pronunciation practice?" Our hypothesis was that Korean university students would appreciate the opportunity to use mobile technology to enhance their English study. We believed students would use their smartphones to practice and that this usage would be perceived as easy and useful.

This first hypothesis was largely confirmed. A total of 210 out of 211 participants confirmed that they had used the ASR on their smartphones to study during the study period (although the precise amount of study is not determined). Table 1 displays the results of the quantitative items testing this research question (questions 2, 3, and 4 on the questionnaire). All items showed a positive result, with a net agreement over 70% on each. The highest positive result was found on the item "Smartphone practice is a useful method to study

pronunciation" ($M = 4.17, SD = 0.80$). High positive results were also found on the items "It was easy to practice using my smartphone" ($M = 4.11, SD = 0.99$) and "I felt the smartphone practice helped me improve" ($M = 4.09, SD = 0.87$). This last item also showed the highest level of net agreement among participants.

This finding was further confirmed by the results of the open-ended response on the survey instrument. Out of 211 total participants, 104 left open-ended comments. Of these, 22 out of 104 (21.15%) identified the whole process as "good," "fun," or "interesting," (i.e., "It was good") while 17 out of 104 (16.35%) identified it as "helpful" or "useful" for improving their English (i.e., "Helpful for me, after the class I can say English without worries").

TABLE 1

Response to Questions on Learner Response to ASR

Item	M	SD	% Net Agree
Q2. It was easy to practice using my smartphone.	4.11	0.99	71.8
Q3. I felt the smartphone practice helped me improve.	4.09	0.87	75.0
Q4. Smartphone practice is a useful method to study pronunciation.	4.71	0.80	72.4

*Ratings are on a 5 point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

4.1.2. Which type of application would participants consider the most useful?

The second research question was: "Which type of application would participants consider the most useful?" Our hypothesis was that Korean university students would prefer to use translation applications for pronunciation study and that they would show a marked preference for Korea-based applications in doing so. The reason for this hypothesis is the relative popularity of Naver and its English/Korean translation application "Papago" among university students. This research question was investigated by a single open-ended response on the research instrument.

This second hypothesis was confirmed, although not to the degree anticipated. The majority of participants (78.4%) did report using a translation application for their period of self-study. Papago and Naver were the preferred choice of a plurality of participants (45.5%), with the next most-popular choice being Google Translate (28.4%) or a combination of Google and Naver (4.5%). The results are given in Table 2.

TABLE 2
I Practiced Using This Smartphone App

Application Name	Number	% of Total
Papago/Naver	79	45.14
Google	50	28.57
Papago + Google	8	4.57
Others	38	21.71

Note. Others include Kakaotalk, Siri, Zoom, Amazon Echo, Voice Recorder, and several iPhone applications.

The open-ended response results also confirmed the hypothesis. Out of 104 open-ended responses, 5 (4.81%) specifically mentioned the advantage of studying using translation applications (i.e., "Practicing using Papago was fun"). Several participants (4, 3.85%) referenced the advantage of using technology for language study without mentioning any specific application (i.e., "Great to use up to date learning service"). There were no reports in the open-ended response of problems using an application or hesitancy regarding the ASR technology itself during the study period.

4.1.3. How do learners feel about using ASR as a testing method?

The third research question was: "How do learners feel about using ASR as a testing method?" Our hypothesis was that participants would be accepting of ASR for testing with some hesitation among those who experienced technical problems. There were several assumptions to this hypothesis: 1) Participants would be interested in taking a test in an unusual and unfamiliar format; 2) Participants would perform better than expected due to their time practicing the material; and 3) Participants who experienced technical difficulties during testing would report greater hesitancy toward the testing method.

This hypothesis was more tenuously confirmed than the previous two. Table 3 displays the results of the quantitative items testing this research question. Participants believed the system was fair, showing a high level of agreement with the statement "The teacher's organizing system was fair." ($M = 4.3$, $SD = 0.88$) This item was explained as the way the teacher explained the rules, ran the test, provided feedback about using the equipment correctly and allowed a reasonable opportunity for retries. Responses to the test were also largely positive, with a high level of agreement with the statement "I was comfortable taking the pronunciation test" ($M = 4.2$, $SD = 0.91$) and "I preferred speaking the pronunciation test to taking a written test" ($M = 4.07$, $SD = 0.98$). However, participants were more divided on the question of whether the test itself was fair or how they had performed on the test. The lowest mean score and highest variance on any quantitative item came on the statement "I felt I did well on the pronunciation test" ($M = 3.8$, $SD = 1.06$). Similar scores were found on

the item "The computer listened to everyone the same way" ($M = 3.9$, $SD = 0.98$). The open-ended response provided further insight into the quantitative results. Out of 29 total responses which specifically referenced the testing method, 20 (19.23% of 104 total responses) identified the participant as "happy" or "satisfied" with the testing method or results. Of the 9 other responses, 4 reported technical problems with the test (i.e., "It's not a bad method, but the microphone didn't work well") while the remaining 5 indicated a general hesitancy with the test itself (i.e., "I did not do well in the test, but I learned the correct pronunciation of some words").

TABLE 3

Response to Questions Concerning Using ASR as a Testing Method

Item	*M*	*SD*	% Net Agree
Q5. I preferred speaking the pronunciation test to taking a written test.	4.07	0.98	72.4
Q6. I was comfortable taking the pronunciation test.	4.20	0.91	75.7
Q7. I felt I did well on the pronunciation test.	3.80	1.06	62.2
Q8. The computer listened to everyone the same way.	3.90	0.98	75.7
Q9. The teacher's organizing system was fair.	4.30	0.88	81.4

*Ratings are on a 5 point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

4.1.4. Open-ended response

The questionnaire included one item allowing open-ended commenting with the request: "My comment." Out of a total of 211 participants, 104 responded with comments in a total of 736 words. These comments ranged from one-word responses (i.e., "good") to multiple-sentence responses covering several themes. Using the constant comparative method, these responses were thematically coded and categorized in two different ways (Maykut & Morehouse, 1994). First, comments were coded according to their response to the three main research questions: "What responses do learners have to using ASR for pronunciation practice," "Which type of application would participants consider the most useful," and "How do learners feel about using ASR as a testing method." Responses which specifically referenced the test (i.e., "Happy and satisfied with the test," "the quiz was interesting") were coded as responding to the third question. Responses specifically referencing the use of a particular application (i.e., "practicing using Papago was fun," "the Google app is accurate") were coded as responding to the second question. Responses which made a general comment on the total experience without reference to the test or learning application (i.e., "It was good," "Thank you") were coded as responding to the first question. These results are collected in Table 4.

The results were also coded according to the categories and themes of the Technology

Acceptance Model (TAM) as catalogued in Kemp, Palmer, and Strelan (2019). Specifically, the categories of Perceived Usefulness (PU), Perceived Ease of Use (PEOU) and Attitude (AT) were used. Note that while Attitude is sometimes omitted from TAM models, it is still retained due to its usefulness in certain circumstances, particularly in a voluntary setting such as the one used in this experiment (Kemp et al., 2019). These results are collected in Table 5.

TABLE 4
Qualitative Responses to Research Questions

Theme	Number	% of Total
Good, fun, or interesting experience	22	21.2
Helpful or useful experience	17	16.3
Referenced using specific translation app	5	4.8
Referenced using ASR technology generally	4	3.8
Reported satisfaction with the testing method	29	27.9
Reported technical problems with the test	4	3.8
Reported hesitancy with regard to the test	5	4.8

Note. Totals out of 104 responses. Some comments contained multiple themes

TABLE 5
Qualitative Responses Coded According to the TAM

Category	Number	% of Total
PU (positive response)	26	25.0
PU (negative response)	1	0.1
PEOU (positive response)	8	7.7
PEOU (negative response)	2	0.2
AT (total)	67	64.4
AT (positive)	56	53.8
AT negative or neutral)	11	10.6

Note. All 104 responses coded.

4.2. Discussion

This study looked at the convergence of two elements of English language teaching: the need for methods of teaching pronunciation to English language learners and the trend toward incorporating technology into the English classroom. A Korean university setting offers a unique opportunity to study this convergence. University students in Korea typically must demonstrate some degree of English proficiency but have only limited opportunities for L2 English immersion within which to develop their pronunciation skills (Kim, 2015).

In addition, Korean university students typically have broad access to ASR technologies which could benefit their pronunciation skills and show a high degree of openness to incorporating technology into their learning modes.

Regarding the need for methods of teaching pronunciation, this study is encouraging in the incorporation of technology in the classroom. Following the TAM coding shown in Table 5, it can be suggested that many Korean students are open to the incorporation of technologies such as ASR in the classroom. This is particularly beneficial in a Korean English-learning context, where learners exhibit a much higher than average degree of fossilization in their English skills (Kim, 2015). Furthermore, Isaacs (2017) has argued that test-takers in classroom settings might prefer speaking tests of any kind to be administered by their teachers. The findings of this study provide further support for his hypothesis—particularly the high level of satisfaction reported with the testing despite certain technical problems (see Table 4).

As for the trend toward incorporating technology in the classroom, the high degree of positive responses recorded in Tables 4 and 5 as well as the generally positive responses found on the quantitative elements of the questionnaire suggest an environment within which teachers will find willing recipients for pedagogical research integrating technology (Ducar & Schocket, 2018). Surprisingly, students were initially unaware of the possibility of incorporating smartphone ASR capabilities into an English learning context, but after seeing the benefits of such incorporation demonstrated they seemed open to the idea of pursuing this further.

The researchers believe there is significant potential in incorporating Automated Speech Response technology into the English classroom. One interesting possibility concerns the de-centering of the teacher as students' focus point. Once the use of the technology is demonstrated, students may be free to use the technology with little additional input from the teacher. Students may come to rely less on the teacher and more on the practice with their own devices. Another possibility is that Automated Speech Response practice may help overcome issues with performance shyness, as students may be more willing to practice freely without fear of error. A third possibility is that Automated Speech Response practice may help move students away from a "native speaker" focus on pronunciation and toward a more communicative approach. With a technological device rather than a human being, the user's focus may shift toward being understood clearly rather than speaking exactly as a native speaker would speak. Any one of these possibilities could be further explored in the classroom by enterprising educators.

This study did, however, confirm that earlier reported imperfections with ASR technology are still in place—making this method one of which researchers and teachers alike must remain cautious. One example noted by researchers during testing was that the ASR system used for the test still produced errors with certain proper nouns. For example, the sentence

"Tell me about John's party last week" was transcribed variously by the system as 'Jones', 'Joe's' or 'John's' despite no recognizable differences in pronunciation by students.

5. CONCLUSION

Mobile assisted language learning (MALL) has become the most recent and most widely accessible iteration of what used to be called "computer assisted language learning." As smartphone use has proliferated, and the ASR technologies integrated within smartphones has become increasingly more sophisticated, it behooves educators to explore the possibilities of integrating MALL in the English classroom. Recent studies into mobile device ownership and technology show that learners expect not just to be able to use their devices but also to receive support from teachers of their adoption of such educational resources (Hoi & Mu, 2021). This study represents an attempt to integrate the interests of teachers and students through a partially-guided use of ASR technology in the classroom.

This study has several key limitations which future research could address. These limitations fall into two broad categories: student use of ASR (and other technologies) and the effectiveness of ASR itself when it comes to improving English speaking skills. This study focused on student use of ASR over the course of one university semester and did not attempt to assess whether students would continue using this in the future or to what degree students had used the technology during the semester. As this study focused on student response to ASR rather than the actual effectiveness of the technology itself, it offers no insight into the pedagogical benefits it may bring. The positive response of students to the use of ASR in this study, however, does provide a hopeful note that, should the technology be effective, students would likely adopt it with some instructor guidance.

Concerning the first limitation, future studies should ask participants to assess their daily or weekly use of ASR to study pronunciation. They could also ask students to estimate their likelihood of continuing to use the technology in the future. They could interview some students who wanted to discuss their use of ASR in more depth. Without prompting, participants in the current study showed some interest in incorporating this kind of technology into their future English studies. As for the second limitation, research so far has been tentatively positive. As technological capabilities improve, ASR may prove to be a vital part of the English language classroom. One interesting point of consideration for future research would be the degree to which English proficiency influences perceptions of Automated Speech Response for pronunciation learning. Do students perceive it as more useful if their levels are low or stagnating? Or does perception of utility increase along with English proficiency? Such questions were not examined in this research but could be worth investigating.

This research began with the idea of integrating already-existing ASR technology into the classroom. Such integration offers opportunities not only for learning pronunciation but for developing English skills of any kind using technology that is already at the students' fingertips. In keeping with previous studies, the student response to this suggests that teachers who choose to experiment in this direction will find willing participants. It is up to teachers and researchers to find methods of use and instruction which have real and lasting pedagogical value.

Applicable level: Tertiary

REFERENCES

Abdel-Reheem Amin, E. (2020). A review of research into Google Apps in the process of English language learning and teaching. *Arab World English Journal (AWEJ)*, *11*(1).

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.

Couper, G. (2017). Teacher cognition of pronunciation teaching: Teachers' concerns and issues. *TESOL Quarterly*, *51*(4), 820-843.

Cucchiarini, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, *51*(10), 853-863.

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

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*, *51*(4), 779-795.

Evanini, K., Hauck, M. C., & Hakuta, K. (2017). Approaches to automated scoring of speaking for K–12 English language proficiency assessments. (Research Report No. RR-17-18). Princeton, NJ: Educational Testing Service.

Evers, K., & Chen, S. (2021). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, *59*(4), 669-685.

Garcia, C., Kolat, M., & Morgan, T. (2018). Self-correction of second language pronunciation via online, real-time, visual feedback. *In Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference* (pp. 54-65). Ames, IA: Iowa State University.

Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: A user perception study. *Assessment in Education: Principles, Policy & Practice*, *28*(1), 58-76.

Guskaroska, A. (2019). ASR as a tool for providing feedback for vowel pronunciation practice (Doctoral dissertation, Iowa State University).

Harding, L. (2011). *Accent and listening assessment: A validation study of the use of speakers with L2 accents on an academic English listening test*. Peter Lang Publishing, Frankfurt. Germany.

Hoi, V. N., & Mu, G. M. (2021). Perceived teacher support and students' acceptance of mobile-assisted language learning: Evidence from Vietnamese higher education context. *British Journal of Educational Technology*, *52*(2), 879-898.

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.

Isaacs, T. (2017). Fully automated speaking assessment: Changes to proficiency testing and the role of pronunciation. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 570-584). Abingdon, UK: Routledge.

Kemp, A., Palmer, E., & Strelan, P. (2019). A taxonomy of factors affecting attitudes towards educational technologies for use with technology acceptance models. *British Journal of Educational Technology*, *50*(5), 2394-2413.

Kim, C. (2015). Demotivational factors influencing compulsory English language learning of Korean university students in an EFL context. Unpublished doctoral dissertation, Korea University, South Korea

Kim, I. S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Journal of Educational Technology & Society*, *9*(1), 322-334.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369-377.

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

Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *Languages*, *2*(3), 11.

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.

Maykut, P., & Morehouse, R. (1994). *Maykut, Pamela, and Richard Morehouse, Beginning Qualitative Research: A Philosophic and Practical Guide*. Bristol, PA: Falmer Press, 1994.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57*, 25-42.

McCrocklin, S. (2018). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal*, *36*(2), 119-137.

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

McCrocklin, S. (2019b). Dictation programs for second language pronunciation learning: Perceptions of the transcript, strategy use and improvement. *Konińskie Studia Językowe*, *7*(2), 137-157.

McCrocklin, S., Humaidan, A., & Edalatishams, E. (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). Iowa State University.

McCrocklin, S., & Edalatishams, I. (2020). Revisiting popular speech recognition software for ESL speech. *TESOL Quarterly*, *54*(4), 1086-1097.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*(4), 418-440.

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

Mroz, A. (2020). Aiming for advanced intelligibility and proficiency using mobile ASR. *Journal of Second Language Pronunciation*, *6*(1), 12-38.

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, *20*(2), 225-243.

Sagita, M., Jamaliah, J., & Balqis, N. (2021). Students' perception about Google translation tool in learning English. *Getsempena English Education Journal, 8*(1), 26-40.

Shadiev, R., Hwang, W. Y., Huang, Y. M., & Liu, C. J. (2016). Investigating applications of speech-to-text recognition technology for a face-to-face seminar to assist learning of non-native English-speaking participants. *Technology, Pedagogy and Education, 25*(1), 119-134.

Wang, Y. H., & Young, S. C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, *31*(6), 493-504.

Weng, F., Yang, R. J., Ho, H. J., & Su, H. M. (2018). A TAM-based study of the attitude towards use intention of multimedia among school teachers. *Applied System Innovation*, *1*(3), 36. https://doi.org/10.3390/asi1030036

Yadav, M. S. (2021). Role of Social Media in English Language Learning to the Adult Learners. *International Journal of Linguistics, Literature and Translation, 4*(1), 238-247.

Zanzotto, F. M. (2019). Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, *64*, 243-252.

Zechner, K., Chen, L., Davis, L., Evanini, K., Lee, C. M., Leong, C. W., Wang, X., & Yoon, S.-Y. (2015), Automated scoring of speaking tasks in the test of English-for-teaching (*TEFT™*). (Research Report No. RR-15-31) Princeton, NJ: Educational Testing Service.

Zheng, J., & Li, S. (2020). What drives students' intention to use tablet computers: An extended technology acceptance model. *International Journal of Educational Research*, *102*, (Article No. 101612). https://doi.org/10.1016/j.ijer.2020.101612

APPENDIX A

TABLE 6
Questionnaire Items

N	Question
1	I practiced using this smartphone app: _____
2	It was easy to practice using my smartphone.
3	I felt the smartphone practice helped me improve.
4	Smartphone practice is a useful method to study pronunciation.
5	I preferred speaking the pronunciation test to taking a written test.
6	I was comfortable taking the pronunciation test.
7	I felt I did well on the pronunciation test.
8	The computer listened to everyone the same way.
9	The teacher's organizing system was fair.
10	My comments: _____

APPENDIX B

TABLE 7
Full Result of Questionnaire Items

Item	*M*	*SD*	Strongly Disagree %	Disagree %	Neutral %	Agree %	Strongly Agree %	Net %
Q2	4.11	0.99	1.4	4.3	22.5	24.9	46.9	71.8
Q3	4.09	0.87	0.5	2.9	21.6	37.0	38.0	75.0
Q4	4.17	0.80	0.5	1.0	21.4	30.0	42.4	72.4
Q5	4.07	0.98	1.4	4.8	21.4	30.0	42.4	72.4
Q6	4.20	0.91	0.5	3.3	20.5	27.1	48.6	75.7
Q7	3.80	1.06	1.9	10.0	25.8	29.7	32.5	62.2
Q8	3.90	0.98	1.9	5.7	20.5	27.1	48.6	75.7
Q9	4.30	0.88	1.9	0.0	16.7	27.1	54.3	81.4

Note. Question text shown in Appendix A
[*]Ratings are on a 5 point Likert scale from 1 (strongly disagree) to 5 (strongly agree).