2019

ASR as a tool for providing feedback for vowel pronunciation practice

Agata Guskaroska
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Bilingual, Multilingual, and Multicultural Education Commons, and the English Language and Literature Commons

Recommended Citation
Guskaroska, Agata, "ASR as a tool for providing feedback for vowel pronunciation practice" (2019). Graduate Theses and Dissertations. 17020.
https://lib.dr.iastate.edu/etd/17020

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
ASR as a tool for providing feedback for vowel pronunciation practice

by

Agata Guskaroska

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Major: Teaching English as a Second Language/Applied Linguistics

(Computer Assisted Language Learning)

Program of Study Committee:
John Levis, Major Professor
Volker Hegelheimer
Charles Nagle

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2019

Copyright © Agata Guskaroska, 2019. All rights reserved.
DEDICATION

In dedication to my mother, who has always been my greatest supporter. *La vita e breve, utile bouna.*
TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................ v
LIST OF TABLES ............................................................................................................. vi
NOMENCLATURE ......................................................................................................... vii
ACKNOWLEDGMENTS ............................................................................................... viii
ABSTRACT ....................................................................................................................... ix

CHAPTER 1. INTRODUCTION ....................................................................................... 1

CHAPTER 2. LITERATURE REVIEW ............................................................................ 5
  ASR and Pronunciation ................................................................................................. 5
  Pronunciation improvement ..................................................................................... 6
  Accuracy for providing feedback ............................................................................. 8
  Learners’ Benefits: Autonomy, Awareness Raising and Attitudes Towards ASR .................................................................................................................. 10
  A Comparison between Macedonian and English Vowel Systems ..................... 12
  The Study ..................................................................................................................... 15
  Research Questions ................................................................................................ 16

CHAPTER 3. METHODS ................................................................................................ 18
  Participants .................................................................................................................. 19
  Task ............................................................................................................................. 21
  Procedure ..................................................................................................................... 24
  IRB approval .......................................................................................................... 27
  Data Analysis ............................................................................................................... 27

CHAPTER 4. RESULTS .................................................................................................. 30
  Vowel pronunciation improvement ............................................................................. 30
  ASR Accuracy for vowel pronunciation practice ........................................................ 33
  Learners attitudes and beliefs towards ASR ............................................................... 38

CHAPTER 5. DISCUSSION AND CONCLUSION ....................................................... 41
  Discussion ..................................................................................................................... 41
  Limitations and directions for future research .............................................................. 50
  Implications ............................................................................................................. 51
  Conclusions ................................................................................................................. 52

REFERENCES ................................................................................................................. 54

APPENDIX A. LIST OF MINIMAL PAIRS FOR ANALYSIS ..................................... 60
LIST OF FIGURES

Figure 1. Macedonian and English vowel diagrams (adapted from Kirkova-Naskova, 2012) ........................................................................................................ 14
Figure 2. Summary of Research Methods ........................................................................ 18
Figure 3. The phases of the procedure ........................................................................ 27
Figure 4. Listeners’ accuracy scores for NNSC and EXP (M) ........................................ 31
Figure 5. ASR and Native listeners’ recognition scores ................................................ 36
Figure 6. Recognition scores for individual vowels ...................................................... 37
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.</td>
<td>Linguistic Background Questionnaire (EXP)</td>
<td>20</td>
</tr>
<tr>
<td>Table 2.</td>
<td>Linguistic Background Questionnaire (NNSC)</td>
<td>21</td>
</tr>
<tr>
<td>Table 3.</td>
<td>Overall accuracy scores on pre-test and post-test (M and SD)</td>
<td>30</td>
</tr>
<tr>
<td>Table 4.</td>
<td>Learners’ pronunciation improvement per individual vowels (NNSC)</td>
<td>32</td>
</tr>
<tr>
<td>Table 5.</td>
<td>Learners’ pronunciation improvement per individual vowels (EXP)</td>
<td>32</td>
</tr>
<tr>
<td>Table 6.</td>
<td>Mean Recognition Scores</td>
<td>34</td>
</tr>
<tr>
<td>Table 7.</td>
<td>ASR and Native listeners’ recognition scores</td>
<td>35</td>
</tr>
<tr>
<td>Table 8.</td>
<td>ASR Recognition scores for NNS and NS</td>
<td>37</td>
</tr>
</tbody>
</table>
## NOMENCLATURE

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR</td>
<td>Automated speech recognition</td>
</tr>
<tr>
<td>CALL</td>
<td>Computer-assisted language learning</td>
</tr>
<tr>
<td>CAPT</td>
<td>Computer-assisted pronunciation training</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a Foreign Language</td>
</tr>
<tr>
<td>ESL</td>
<td>English as a Second Language</td>
</tr>
<tr>
<td>L1</td>
<td>First language (native language)</td>
</tr>
<tr>
<td>L2</td>
<td>Second language</td>
</tr>
<tr>
<td>NS</td>
<td>Native speakers</td>
</tr>
<tr>
<td>NNS</td>
<td>Non-native speakers</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SLA</td>
<td>Second language acquisition</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. John Levis, and my committee members, Dr. Volker Hegelheimer and Dr. Charles Nagle, for their guidance and support throughout the course of this research. In particular, I would like to thank John for his valuable feedback and observations on this topic.

In addition, I would also like to thank my friends, colleagues, and the department faculty and staff for making my time at Iowa State University a wonderful experience. I would like to show gratitude to the members of the PRG group and in particular, my friend Tim Kochem for his feedback in preparation for the presentation. I would also like to thank the English department at ISU for supporting this study by TESL/ALT Small Research Grant. I want to also offer my appreciation to my mother, my husband and son, my friends Sanja Baloska Stojanovska and Snezhana Karadzova Popova for their assistance with obtaining the recordings, and those who were willing to participate in my study, without whom this thesis would not have been possible.
ABSTRACT

The purpose of the study is to examine the usefulness of mobile-assisted ASR dictation systems (Gboard, Siri or voice dictation on smartphones) for vowel pronunciation practice by looking at three aspects of its usefulness: pronunciation improvement by using ASR, accuracy of recognition, and the learners’ attitudes towards using this system. A list of 30 words containing minimal pairs of four contrasts /i/, /ɪ/; /æ/, /ɛ/; /u/, /ʊ/; /ɑ/, /ʌ/ and some distractors were given to 21 Macedonian EFL learners, divided into two groups, an experimental (n=11) and a control group (n=10). A mixed methods approach was used in this study. The quantitative part of the study included pre-test and post-test recordings which were transcribed by 10 native listeners to measure their accuracy gains, as well as a comparison between ASR written output of native speakers and that of non-native speakers, and another comparison between ASR written output of non-native speakers and human judgments. The qualitative analysis explored learners’ attitudes towards ASR by analyzing students’ Facebook posts throughout the practice period.

Findings showed that the experimental group improved their accuracy while the control group did not show any improvements. Next, the findings demonstrated a close relationship between ASR written output and human judgment with an acceptable agreement for most of the vowels. Nonetheless, ASR did not show high recognition of native speech, especially for the vowels /o/ and /ʌ/. Qualitatively, the learners’ Facebook posts showed positive attitudes towards ASR. An occasional frustration with inaccurate feedback was also reported but learners generally enjoyed the training and found ASR to be practical and a safe environment for practice. This study recommends inclusion of
mobile-assisted ASR in the EFL classrooms for raising students’ awareness of the vowel sounds in the English language with careful guidance from the teacher as well as focused and structured practice using individual words.
CHAPTER 1. INTRODUCTION

Improved pronunciation is important in learner progress towards spoken proficiency (Huensch, 2019). Despite its importance, pronunciation is often a neglected area in the second language (L2) classroom due to lack of time and the difficulties the teachers face in providing individual feedback to each learner (Isaacs, 2009; Kelly, 1969; McCrocklin, 2016). Besides lack of time, the teachers may encounter lack of resources, lack of confidence in teaching pronunciation, and uncertainty of how to integrate pronunciation into the curriculum (Levis & Grant, 2003).

Integrating pronunciation instruction in a regular ESL/EFL course can be challenging, but it is very important, especially in an EFL setting where the exposure to native language can be limited (McCrocklin, 2014). Various techniques can be used in these classrooms, such as utilization of known words, explanation of how to produce a sound, communicative activities, modeling and individual correction, reading aloud, self-monitoring and self-correction, and computer-assisted language learning (Nikbakht, 2010). While receiving feedback is essential for success, learners may often feel intimidated by explicit correction of their speech in front of their peers which may lead to hesitation to speak in a foreign language. Hence, EFL teachers may struggle to find the right approaches and techniques that can work in their classroom, bearing in mind that classes may have limited resources and a large number of students, and that (non-native) teachers may feel less confident in providing themselves as models.

With the growing emergence of computer-assisted pronunciation training (CAPT) tools, researchers have investigated possibilities to assist the process of learning pronunciation and their integration into EFL/ESL classrooms (McCrocklin, 2019; Mroz,
While there are many available CAPT tools, feedback is what is mostly lacking with CAPT (Levis, 2007), and research of tools, systems, and software such as Automated Speech Recognition (ASR) is needed in the field. This last option, ASR, has particularly strong promise as a tool for pronunciation. Levis and Suvorov (2014) define ASR as “an independent, machine-based process of decoding and transcribing oral speech” (p. 1) which turns the speech signal into text. While ASR can be used for several purposes, such as business dictation and special needs accessibility (Kim, 2006), its use for language learning has considerably increased throughout the years (Ahn & Lee, 2016; Wang & Young, 2015; Witt & Young, 2000).

The potential of ASR systems to identify pronunciation problems and give accurate feedback to learners has been of interest for many researchers (Cucchiarini, Stirk & Boves, 2000a; 2000b; 2002; Doremalen et al., 2016; Eskenazi, 1999a, 1999b). They have explored the idea of using ASR written output as feedback for learners by using commercial dictation systems and/or building ASR-based CAPT systems that are specifically designed for pronunciation practice (Ashwell & Elam, 2017; Coniam, 1998; McCrocklin, 2016; Neri et al., 2008). ASR may offer possibilities for feedback for L2 learners (Derwing, Munro & Carbonaro, 2000), facilitate autonomy (McCrocklin, 2016), create a safe environment for self-practice, save time, promote self-monitoring (Wallace, 2016) and raise students’ awareness of their own intelligibility as well as awareness of “what is at stake in making interactions with others intelligible” (Mroz, 2018, p. 12). ASR has tremendous potential and looks promising for pronunciation self-access work by providing a safe, low-anxiety environment for learners (Chen, 2011).
Despite all the benefits that it can offer, the accuracy of ASR for language learning applications remains uncertain. Clearly, the goal of ASR in identifying mispronunciations is to reach the golden standard of human raters. Even though many studies have shown that this standard is out of reach (Derwing et al., 2000; Kim, 2006), Coniam (1998) found that the system got higher recognition on smaller units, that is a single word. Levis and Suvorov (2014) also argued that ASR is more accurate for phone-level recognition than for prosody. On the other hand, studies found that human ratings of learners’ speech were closely related to ASR feedback, which is in favor of the use of these systems (Cucchiarini et al., 2000a, 2000b, 2002; Gretter, Allgaier, Tchistiakova, & Falavigna, 2019; Witt & Young, 2000). Considering the constantly advancing development of technology, more research is needed to evaluate the recent, updated ASR systems.

As discussed above, research has examined the effectiveness of ASR-based CAPT systems specifically designed for pronunciation practice and commercial dictation systems as tools to help pronunciation acquisition. While developing ASR-based systems for pronunciation practice is invaluable, many EFL teachers lack the skills to develop such systems as well as access to already designed systems of this type. In that regard, using commercial dictation systems which are easily and freely available may find a broader use by EFL teachers around the globe, if the tools are sufficiently accurate. Instead of providing explicit feedback, the way these systems can be used for providing pronunciation feedback is to examine the dictation text (the program’s written output) to see where possible mispronunciation caused the program to transcribe the intended text (McCrocklin, 2019). Smartphone ownership is rapidly growing worldwide (Pew Research Center, 2019) and recent studies on this topic are in favor of mobile-assisted ASR dictation systems (Liakin,
Cardoso & Liakina, 2015; Mroz, 2018). From a practical point of view, these systems may find their use in classrooms around the world.

Small countries, such as Macedonia, present an interesting EFL setting for exploration. English is introduced at the age of six and is taught throughout the entire primary and secondary education in Macedonia. Nonetheless, because of large classes (usually ranging from 20 to 36 students), lack of teacher preparation, and limited resources, very often leading to neglecting pronunciation in the regular EFL classes, an exploration of practical tools is highly needed in these types of settings.

Even though a few studies have explored pronunciation practice using ASR dictation systems on smartphones, the topic is underresearched and needs more exploration to discover the potential of these tools. Therefore, this study aims to explore the potential of ASR dictation systems (Gboard – the Google keyboards, Siri and voice dictation programs available on smartphones) for improving vowel pronunciation production in an EFL setting. The main purpose of the study is to explore 1) the learners’ level of improvement of vowel production after using ASR for pronunciation practice’, 2) the degree of accuracy of these systems’, and 3) the learners’ attitudes towards the systems. By examining these three aspects of ASR usefulness, overall conclusions will be drawn about the potential of these systems for vowel pronunciation practice in an EFL setting. ASR will be examined by allowing word-level practice and focusing only on vowel sounds that, based on a comparison between Macedonian and English vowel systems, are selected as problematic for Macedonian learners.
CHAPTER 2. LITERATURE REVIEW

The study was framed using two bodies of research: previous studies on the use of ASR for pronunciation improvement and comparisons of Macedonian and English vowel systems. The first section synthesizes previous work that has been done in the area of ASR and pronunciation with the aim to provide insight into the approaches, methods, and findings of these studies and to identify a gap which needs further exploration. The literature is organized around three main aspects which are important for the evaluation of this software: pronunciation improvement; accuracy for providing feedback and learners’ attitudes towards ASR. The second section provides a comparison between the Macedonian and the English vowel systems with the aim to identify sounds that might be problematic for these learners. Finally, after reviewing these two areas of research, the focus and the research questions of the current study are presented.

ASR and Pronunciation

The purpose of this section is to provide an overview of research on ASR and its usefulness, ability and potential to serve as a tool for pronunciation improvement, and to identify a gap that needs further exploration. The studies are organized according to the following aspects of ASR usefulness:

a) Pronunciation improvement (how much the participants improved their pronunciation by using ASR);

b) Accuracy of feedback (how accurate ASR systems are for recognizing native and non-native speech, and providing feedback to the learners).

c) Learners’ attitudes towards ASR: Satisfaction with using ASR for improving their pronunciation skills, learner autonomy (the extent to which ASR facilitated learner
autonomy); raising learners’ awareness (how effective ASR was for raising learners’ awareness regarding their pronunciation).

**Pronunciation improvement**

The ultimate goal of using various techniques, strategies and approaches in the integration of pronunciation in the classroom, is to improve the learners’ pronunciation skills in perception and production. The research reports on various ways of evaluating ASR’s usefulness for pronunciation instruction (such as learners’ attitudes and beliefs, fostering autonomy, accuracy of the system) all of which are relevant in evaluating the system, but the main purpose, after all, is to see improvement in the learners’ speech. Most of the studies in the field reported that the participants improved their pronunciation (McCrocklin, 2019; Cucchiarini, Neri, & Strik, 2009) and/or benefited by using ASR. Nonetheless, most researchers acknowledge that pronunciation ‘slightly’ improved and it may have been influenced by other factors.

The exploration of pronunciation improvement of learners who used ASR for pronunciation training, most commonly compared to another control group who did not receive the same training, mostly showed positive results. A few studies, (such as Cucchiarini et al., 2009; McCrocklin, 2019; Neri et al., 2008; Wang & Young, 2015) investigated ASR-based CAPT systems by making a comparison of learners’ speech before and after an ASR training period. Cucchiarini et al. (2009) explored the ASR system *Dutch-CAPT* with immigrants learning Dutch and found that the experimental group improved the most even though the improvement was just slightly higher compared to the other groups. The authors also noted that the greatest improvement of this group might be due to the lowest level of these learners before the training. Another study in line with these findings is Neri et al., (2008) which explores the improvement of segmental quality in L2 Dutch by using the same
Dutch-CAPT system. Neri et al. found that the results on global segmental quality showed that the group who used ASR training made the greatest mean improvement. Moreover, positive results regarding accuracy gains were found by Wang and Young (2015) who recorded English L2 learners’ speech samples before and after an eight-week training using an ASR-based system. Their findings also showed that learners benefited by using the system.

Researchers are mostly in favor of ASR-based systems for improving pronunciation. Nonetheless, while in the past most studies focused on computer-assisted language learning (CALL), and largely on ASR-based CAPT systems for pronunciation practice, a few studies also explored ASR dictation systems and their potential use in the EFL/ESL classrooms. With the spread of mobile devices, mobile-assisted ASR dictation systems have lately become of interest for researchers in the field (Liakin et al., 2015; Mroz, 2018). The accessibility, familiarity and practicability of mobile phones can offer potential advantages for L2 learning such as a self-paced, self-directed learning approach (Victori & Lockhart, 1995). Also, mobile phones can help overcome many of challenges that traditional EFL classrooms face, such as lack of time, lack of language-use opportunities, individual feedback and interaction (Anh & Lee, 2016).

Research on the use of mobile devices for pronunciation teaching and learning is relatively new in the field. Liakin et al. (2015) argue that no previous studies made this kind of investigation before their own study. They stated with the assumption that while mobile devices were considered to be a distraction in the classroom in the past, nowadays their use in the classroom should be investigated. In addition, the results of studies that investigated mobile-assisted dictation ASR tools showed the learners improved after an ASR training
period. Liakin et al. (2015) found that French L2 learners’ pronunciation of the sound /y/ improved after using a commercial ASR application (Nuance Dragon Dictation) for their L2 pronunciation learning practice. Following up on this study, and McCrocklin (2016) who also explored ASR dictation systems, Mroz (2018) also looked at French L2 learners and the possibility of using ASR to raise the students’ awareness of their intelligibility (the way people hear them). Her findings indicated that the learners improved their pronunciation by using Gmail and the French language pack. Clearly, the use of mobile phones in the L2 classroom seems to be beneficial and practical, but its use needs further investigation.

**Accuracy for providing feedback**

An ideal ASR system will identify learners’ errors at the same level as humans’ perception. However, ‘the holy grail of a computer that matches human speech recognition’ is still out of reach (Levis & Suvorov, 2014). Just as humans make mistakes, we can expect errors from ASR systems, but the question is to what extent ASR accuracy is related to human judgments. A method used to evaluate the accuracy of ASR for providing feedback is a comparison between native and non-native speech written output (the extent to which the program recognizes each) and/or a comparison between ASR written output of non-native speech and native speakers ratings or judgments of the same speech. ASR written output is successful if it is similar to native speakers’ judgments.

Early research found lower recognition for non-native speech and concluded that ASR was not accurate enough for pronunciation practice (Derwing et al., 2000; Coniam, 1999). These early researchers pointed out that if ASR programs improved in the future, they may provide a wide range of possibilities. Derwing et al. (2000) used Dragon NaturallySpeaking and found 90% accuracy for native speakers, and 71-73% accurate for very advanced Cantonese and Spanish L1 learners. Native raters successfully transcribed 95-
97.7% of the words but ASR only transcribed 71-73%, which pointed out that the program was far from human judgments. However, another study by Coniam (1998), found that even though Dragon NaturallySpeaking mostly had issues with the recognition of whole t-units (full clauses), it was much more accurate on smaller units, i.e., on a word level. Moreover, other early studies, (LaRocca et al., 1999; Mostow & Aist, 1999) suggested that ASR could be useful only for improvement of segmental features.

Nonetheless, the improvements of accuracy of ASR has led to continuing exploration of the potential of these systems and their application in L2 classrooms. A recent study by Ashwell & Elam (2017) found 89.4% accuracy (Google Web Speech API) for NS and 65.7% for Japanese and Chinese speakers learning English. While many studies have found up to 95% recognition for native speech (Ehsani & Knodt, 1998; Pogue, 2004), the systems’ recognition of non-native speech still remains much lower. If feedback is important for improvement, can commercial ASR still be useful for pronunciation practice? Research on the words which are ‘misheard’ by ASR may show whether this inaccuracy of the system is due to learners’ mispronunciation and whether that could be a useful feedback for the learners’ errors. The current study investigates this issue by exploring the relationship between human judgments and the ASR’s transcription of non-native speech.

Even though a number of limitations of ASR systems exist and the golden standard of human ratings may not be reached, numerous studies show a close relationship between ASR and human judgments (Cucchiarini et al., 2009; Witt & Young, 2000; Neri et al., 2002; O’Brien et al., 2018). In addition, studies show that learners may improve by using less accurate ASR feedback, which suggest that some feedback is better than no feedback. As Neri et al. (2002) point out, through feedback teachers can bring the learners' focus on their
individual problems and motivate them for self-improvement. Hence, while teachers have difficulties finding time to provide feedback, ASR may allow more opportunities for feedback, even if it is not always 100% correct. Additionally, it can be used as a valuable tool in an EFL setting where native-speaking teachers are limited (Kim, 2006) and non-native teachers may feel insufficiently prepared for teaching pronunciation. While the accuracy of ASR-based systems is less than is desirable, research has shown that ASR brings numerous learners' gains and benefits (McCrocklin, 2016; Mroz, 2018).

Learners’ Benefits: Autonomy, Awareness Raising and Attitudes Towards ASR

Another topic that emerges from ASR research is the potential of ASR to facilitate a number of benefits for the learners. Research points out that use of ASR for pronunciation practice is beneficial for the learners in terms of raising students’ awareness of their speech production, facilitating learners’ autonomy and increased motivation.

Raising learners’ awareness of their speech is a valuable aspect towards improving their pronunciation. Many foreign language learners have not had pronunciation training of any kind, or their training is very limited. Raising learners’ consciousness about their speaking by using ASR for practice can allow them to monitor and correct their own errors (Ahn & Lee, 2016), leading to self-directed learning (Victori & Lockhart, 1995). Being intelligible is important for successful communication, and ASR can serve as a way to discover “how people hear you” (Mroz, 2018, p. 1). According to Mroz, the use of ASR raised awareness of one’s own intelligibility and also awareness in what makes speech comprehensible. Even though intelligibility is often put forth as the main goal for pronunciation learning (Levis, 2005), pronunciation accuracy may also assist in improving a person’s social status because strong foreign accents may be stigmatized by native speakers which in some cases may result in social or professional discrimination (Nikbakht, 2010).
The use of ASR in the learning environment can lead towards the establishment of learner autonomy, immediate feedback, input enhancement and multimodal exposure to the forms being acquired (Liakin et al., 2015). McCrocklin (2016) examined students’ autonomous learning beliefs and behaviors after a three-week ASR pronunciation workshop and found significant improvement in the learners’ autonomy beliefs. Furthermore, an exploration of an ASR-based learning system named DISCO to provide pronunciation, morphology, and syntax feedback to learners showed generally positive attitudes towards pronunciation, even though they mentioned a few flaws (Doremalen, Boves, Colpaert, Cucchiarini, & Strik, 2016). The authors speculated that the system is expected to increase students’ motivation, their autonomous work from home and provide wide possibilities for practice. Moreover, learners tend to enjoy these systems because they can produce more output in a low-anxiety environment (Chen, 2011).

Most importantly, learners appreciate the use of ASR. Previous studies reported that, in general, the learners’ attitudes towards ASR systems were positive (Ahn & Lee, 2016; Chen, 2011; Chiu, Liou, & Yeh, 2007). The learners generally believe in the usefulness of the ASR-facilitated training and enjoy its use (Cucchiarini et al., 2009). The wider opportunities for learning and practice of speech may increase motivation and ultimately lead towards speaking ability improvement (McCrocklin, 2014). In Mroz (2018), the students reported satisfaction with the ASR experience, emphasizing that the written output of the program was good feedback for them as it provided a visual representation of their words and potential problems with their pronunciation. Learners’ attitudes, beliefs, and perception of the ASR tool is important in evaluating a learning environment. If the learners enjoy using ASR-
based tools and can see the usefulness of their work, even if the system is not flawless, the learners may improve.

Some researchers have acknowledged occasional students’ frustration when receiving incorrect feedback (Kim, 2006; McCrocklin, 2014) but despite these drawbacks, the overall results of the studies show that ASR creates a safe space for learners that allowed them repeated practice and experimentation with the language (McCrocklin, 2016). To expand on the findings from these studies, the current study will also aim to examine learners' experience, their attitudes, and beliefs in the process of using ASR for pronunciation practice.

A Comparison between Macedonian and English Vowel Systems

L1 transfer has long been explored in SLA. It is well known that the L1 influence can be especially evident in speech and that many errors produced by L2 learners can be attributed to the interference of the learners’ L1 built-in phonological representations (Neri et al., 2002). Research has shown that it is more difficult to acquire similar L2 sounds than distant L2 sounds (Major, 2008). Flege’s Speech Learning Model, or SLM (Flege, 1995; 2007) hypothesizes that the closer an L2 sound is to an L1 sound category, the more difficult it will be for the learners to establish a new category for it. Accordingly, learners can assimilate some sounds to an already existing category in the L1. The SLM also proposes that L2 learners, regardless of their age, can create new phonetic categories of sounds.

This study compared the Macedonian and the English vowel systems in order to hypothesize which sounds may be problematic for these learners and should be included in the training. Most of the recent research that analyses the mispronunciation of vowels of Macedonian EFL learners, done by Kirkova-Naskova (2009; 2010; 2012) also provided support to identify the most commonly mistaken vowels by these learners.
The phonetic system of the Macedonian standard language includes 5 vowels: /i/, /e/, /a/, /o/ and /u/, whereas in some varieties of English there are arguably around 12 vowels and 8 diphthongs (Dodd & Mills, 1996). Figure 1 illustrates a comparison between Macedonian and English vowels diagrams. The first chart provides a comparison between Macedonian vowels and English long (tense) vowels; the second chart compares Macedonian vowels and English short (lax) vowels, and the third compares Macedonian vowels with English diphthongs.
Almost every English vowel presents a potential pronunciation problem for Macedonian learners (Kirkova-Naskova, 2012). The biggest issue for Macedonian EFL learners is the likelihood that they will assimilate the English long and short vowels into one category (Macedonian has a vowel that is somewhere in between the English sounds). For example, the sounds /i/-/ɪ/ are often categorized as the Macedonian sound /i/ which falls somewhere between these two sounds. Although there are many possible difficulties for Macedonian learners in pronouncing English vowels, this study will focus solely on a set of vowel contrasts for monophthongs: /i/-/ɪ/; /æ/-/ɛ/; /u/-/ʊ/; and /ɑ/-/ʌ/. The study examines the pronunciation of individual vowels using minimal pairs. The reasons for selecting these vowel contrasts are as follows

1. /i/-/ɪ/ is a very frequent and difficult contrast for Macedonian ESL learners. In the Macedonian language, there is one sound which is somewhere in between these two sounds and the length of the vocal does not play a distinctive role (Kirkova-Naskova, 2009)
2. /æ/ - /e/ is an important contrast for Macedonian EFL/ESL learners. The Macedonian /e/ is the closest with the English /e/. However, even though /æ/ is more distant from the Macedonian /e/, Macedonian L2 learners rarely hear the difference between these two sounds (Kirkova-Naskova, 2009) and rely on the Macedonian /e/ sound instead.

3. /u/ - /ʊ/ is similar to the first contrast in terms of difficulty for perception by Macedonian learners. Even though the contrast between these two sounds is not frequent in English and has low functional load (Munro & Derwing, 2006), Macedonian learners rarely hear the difference between these two sounds.

4. /ɑ/ - /ʌ/ is interesting because neither of these sounds exists in Macedonian but the learners very often substitute them with the Macedonian sounds /o/ or /a/, respectively (Kirkova-Naskova, 2010). Nonetheless, Kirkova-Naskova (2010) argued that this change of vowel quality was mostly noticed by American raters, as opposed to British raters who were not as sensitive for marking this as a mispronunciation. She labeled the substitution of /ɑ/ with the Macedonian /o/ as ‘vowel raising’ and substitution of /ʌ/ with /a/ as ‘vowel lowering’. In addition, these two sounds (/ɑ/- /ʌ/) may be used interchangeably by the learners, for example, pronunciation of ‘cup’ instead of ‘cop’ or vice versa.

The Study

Previous studies on ASR for pronunciation learning have mostly focused on evaluating ASR CALL systems. Research on using mobile devices for pronunciation teaching and learning has very few studies. While research on ASR-based tools specifically designed for speaking and pronunciation learning is of great importance, access to these tools in actual EFL classrooms may be highly limited. On the other hand, a vast number of students own a smartphone, and each of those mobile devices has some type of ASR
dictation program already installed, or such a program can be downloaded for free. Inspired by the lack of research, the increasing use of mobile phones worldwide, as well as their ubiquity, the current study explores mobile-assisted dictation ASR programs that are freely available on smartphones. In particular, this study explores GBoard – the Google Keyboard, Siri and other voice recognition dictation programs. The initial purpose of the study was to explore only GBoard, however, the Institutional Review Board (IRB) suggested providing freedom for the students to choose their own preferred ASR tool in order to protect their privacy. Nonetheless, all the ASR tools needed to be freely available on smartphones, as this study aimed to investigate easily-available ASR programs that can provide a practical approach in the real-life classroom.

While a number of studies have explored pronunciation improvement, the accuracy of ASR, and learners’ attitudes towards the program by looking at one or two of these aspects, to my knowledge, none of the studies evaluated a mobile-assisted ASR taking all three aspects of usefulness into consideration. To provide a broader view on the big picture, this study starts by assessing pronunciation improvement by pre-test and post-test recordings transcribed by native speakers; continues with evaluations of ASR accuracy by comparing human judgment and ASR; and discusses the main themes that arose from a closed Facebook group discussion posts.

In order to avoid ASR’s assumptions based on context, and as pointed out by previous studies that ASR is more reliable on word level (Conian, 1998) this study uses a two-week training period with individual words instead of t-units or sentences.

**Research Questions**

Mobile-assisted ASR looks promising for facilitating vowel pronunciation practice, so I hypothesize that the learners will show improvement after the treatment period; that ASR
will show an acceptable level of accuracy (10-12% difference between ASR and human judgment, and above 90% recognition of native speech); and that learners will have positive experiences with and attitudes towards using the program. This leads to three research questions in the current study.

1. To what extent do learners improve their pronunciation of vowels after using ASR?

2. How accurately does the ASR a) recognize native and non-native speech and b) relate to human raters?

3. What are the learners’ attitudes towards using ASR for pronunciation training?
CHAPTER 3. METHODS

This chapter provides a description of the research methods used in the study. A mixed method approach was used. The quantitative data was used to answer RQs 1 and 2, that is, to evaluate the usefulness of ASR regarding pronunciation improvement of the learners and accuracy of the software for providing feedback. The qualitative data was used to answer RQ3, that is, to provide the students’ attitudes and experiences with the program. Figure 2 summarizes the design of the study.

Figure 2. Summary of Research Methods
Participants

Two types of participants were included in this study: Speakers and Listeners. The speakers consisted of native English speakers (NS) and non-native English speakers (NNS). The NSs served as a control group for testing the accuracy of ASR for transcribing native speech. This group will be referred to as NS control group (NSC). This group consisted of 10 native speakers of American English, four males and six females, their age ranging from 23 to 38. Additional background information about this group of participants was not collected because their role in the study was only to provide the ASR written output of their speech. The participants were sent an Informed consent via email explaining that they will need to provide the ASR written output of their speech of 30 words, 24 of which were the target words, and six were used as distractors. The ASR output was collected in person by the author using Gboard keyboard’s voice-to-text feature on iPhone and saving their output in a Google doc.

The non-native speakers were divided into two groups: Experimental group (EXP) and NNS Control group (NNSC).

The EXP included 11 Macedonian learners of English, eight males and three females (18-20-year-old), studying English for an average of 12.65 years. The participants completed a linguistic background questionnaire via Qualtrics, summarized in Table 1 below. None of the students had lived or visited an English-speaking country and none had prior pronunciation training, but they all had started learning at a very early age (M=6) and had continuously been studying English throughout their education. English as a foreign language is a compulsory school subject in all the primary schools throughout Macedonia and it is introduced in first grade, at the age of 6. English is studied throughout the entire education (primary and secondary schools) and it is a compulsory course in most of the universities,
regardless of the major. Students self-assessed their English knowledge level to be intermediate (n=5) and high level (n=6). All of the learners had studied three other languages during their primary and high school education: Italian, French, and German. Second foreign language (other than English) is also introduced in all primary schools in Macedonia, at the age of 10, and a third (and in some cases fourth) foreign language is very commonly introduced at a high school level.

The NNS control group (NNSC) included 10 Macedonian learners of English, three males and seven females (18-21-year-old). Like the treatment group, the participants had not lived or visited an English-speaking country and had no prior pronunciation training (Table 2). They were enrolled in an English learning course (some of them at school, some at university level). Their self-assessed level was intermediate (n=3) and high level (n=7).

The listeners were 10 American native speakers, four males and six females. All were graduate students in an Applied linguistics department and were familiar with the area of teaching pronunciation, having previously taken a pronunciation-related course.

Table 1.

Linguistic Background Questionnaire (EXP)

<table>
<thead>
<tr>
<th></th>
<th>Male 8</th>
<th>Female 3</th>
<th>TOTAL 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average age</td>
<td>18.62</td>
<td>18.33</td>
<td>18.54</td>
</tr>
<tr>
<td>Average age of start studying English</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Average years of studying English</td>
<td>12.63</td>
<td>12.33</td>
<td>12.54</td>
</tr>
<tr>
<td>Stay in an English-speaking country</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Other languages spoken</td>
<td>Italian</td>
<td>Italian</td>
<td>Italian</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>French</td>
<td>French</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>German</td>
<td>German</td>
</tr>
<tr>
<td>English pronunciation training</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>English proficiency (self-assessment)</td>
<td>Intermediate 5</td>
<td>Intermediate 5</td>
<td>Intermediate 5</td>
</tr>
<tr>
<td></td>
<td>High 3</td>
<td>High 3</td>
<td>High 6</td>
</tr>
</tbody>
</table>
Table 2.

*Linguistic Background Questionnaire (NNSC)*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average age</td>
<td>18.33</td>
<td>20.14</td>
<td>19.60</td>
</tr>
<tr>
<td>Average age of start studying English</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Average years of studying English</td>
<td>12.33</td>
<td>13.86</td>
<td>12.54</td>
</tr>
<tr>
<td>Stay in an English-speaking country</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Other languages spoken</td>
<td>Italian</td>
<td>Italian</td>
<td>Italian</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>French</td>
<td>French</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>German</td>
<td>German</td>
</tr>
<tr>
<td>English pronunciation training</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>English proficiency (self-assessment)</td>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

The task in this study, practicing pronunciation with ASR, was expected to be engaging for the learners because it is easily accessible, provides immediate feedback and may inspire them to discover their own flaws and become more aware of vowel distinctions made in English. The students’ level was appropriate for this task because they were at a point where they seemed confident about their use of English, but most likely were not aware of wrong pronunciations and differences between certain vowel minimal pairs. Their questions and comments throughout the practice in the Facebook group were expected to indicate their interest and motivation.

**Task**

The students’ task consisted of practicing pronunciation of the selected vowels using a vocabulary list of 24 words during a training period of two weeks, 20-30 minutes a day, with the ASR system (*GBoard* – the Google Keyboards; *Siri* or voice search on smartphones). The main reasons for selecting these mobile-assisted tools were that they are
free, practical and easily available. From a practical point of view, many students nowadays own smartphones and in that regards, these tools may be easy to use in a classroom setting. Additionally, having the tool on their phone could allow them autonomous practice at home and/or in everyday communication through chat with their peers. The vocabulary list given to the learners contained all the target vowels for our study and six distractors for the purposes of the initial recordings. Each sound was represented by three words, for example the sound /i/ in the words leave, feel, sheep. The vocabulary consisted of common, well-known words which are introduced at lower levels of ESL/EFL learning, to avoid mispronunciation due to unfamiliarity with the word (Appendix A).

The participants were located in Macedonia. They were given an informed consent via email which outlined the whole procedure of the training. They were explained that their participation in this study would include two recordings, one before and one after a training period of two weeks where they would practice the pronunciation of vowels using an ASR dictation program on their smartphones, as well as, share their experience in a closed Facebook group. I created a Facebook group with all the participants who were a part of EXP and I posted my announcements in the group for easier communication. The social media Facebook was selected because of its popularity among youth in Macedonia. Facebook, as a convenient, familiar environment where young people already spent a lot time, as opposed to a structured survey at the end of the session, was expected a) to provide more relaxed, natural environment for the participants and b) to capture the learners’ attitudes from the beginning of the practice period, throughout the practice and at the end. While there are many advantages of using a social media group for this purpose, the challenges of using this type of data collection are discussed further in the limitation section. Nevertheless, the Institutional
Review Board (IRB) recommended that this part of the training should be optional for the participants, as some of them may not have a Facebook account, and some of them may not want to participate in an online discussion where their posts can be read by other participants. Hence, even though the Facebook group participation was optional and open for discussion, all the participants agreed to participate. I recommended posting at least four times, but part of the participants posted only two or three times. I used the group announcements to remind them to share their experience and opinion about the training. In the first announcement, I expressed gratitude to the students for their first set of recordings and directed them to practice daily using ASR on their smartphones and share their thoughts and experience with the program. In the following posts, I reminded students to contact me if they had any questions or concerns and encouraged them to share their experience with the ASR practice.

The practice lasted for two weeks and the students practiced autonomously on their own, around 20-30 minutes a day. They were explained that they will be involved into vowel pronunciation practice via informed consent sent to their emails. I provided the vocabulary list for initial practice via email, and by posting it in the Facebook group, and also encouraged students to try with similar words, containing the same vowels (for example, word that rhyme). The learners were instructed to start with one or two minimal pairs per session, depending on how difficult the sounds appeared to them, for example to practice the words live/leave, fil/feel, ship/sheep and/or pen/pan, left/laughed, bed/bad. As the practice progressed, they could add additional words containing the same sounds to expand the practice and also to spend more time on the sounds that were more problematic for them. In the last three days of the session they were instructed to include all the words in a session.
The location for task performance was not specified, but I advised participants to perform the task in a quiet place in order to get more reliable results from the ASR program.

**Procedure**

The data consisted of a linguistic background questionnaire; two sets of speech produced by the 21 Macedonian learners (EXP and NNSC), one obtained before the task (pre-test), and the other obtained after the task (post-test); ASR written output from NSC and EXP; and data from Facebook posts and comments in the closed group created for the purposes of this study. This study focused on the following vowel pairs: /i/, /ɪ/; /æ/, /ɛ/; /u/, /ʊ/; and /ɑ/, /ʌ/ examining each one of them individually. The research procedure for data collection included four phases:

1. An online linguistic background questionnaire followed by pre-test (recordings of non-native speech samples – EXP and NNSC);

2. A treatment period (practice using ASR) and Facebook group posts;

3. Post-test (second set of recorded speech samples – EXP and NNSC); and EXP provided ASR written output in the last day of practice

4. Listeners transcribed the recordings from both groups of the pre-test and post-test; and NSC provided ASR written output;

In Phase 1, the non-native speakers (EXP and NNSC groups) provided linguistic background information using Qualtrics (Tables 1 and 2). After that, they completed the Pre-test phase (creating recording - speech samples of vowel minimal pairs. Most of the EXP participants were recorded with iPhone microphone, using normal pace and quiet background. Their English teacher provided her office and gave the option to each student to go and record themselves. The participants could choose if they want to use their phone or
the teacher’s iPhone, or if they want to record themselves at home and send the recording to the teacher. This led to having a few different voice file types, .aac, .amr, .mp4, m4a, but provided flexibility for the learners. The recordings were sent to the researcher through email. The participants in the Experimental group were given a vocabulary list for the initial recordings of 30 words, 6 of which were distractors. The 24 target words contained simple vocabulary with the selected minimal pairs. The list of words is included in the Appendix. The NNSC group was not given any pronunciation instruction, and they only attended their regular EFL lessons. The NNSC participants recorded their speech by themselves and sent the recording to the researcher via email. The purpose of this was to make a comparison between students who practiced with ASR and students who were in a regular classroom environment.

Phase 2 was the period of practice for the Experimental group (EXP). I informed the students (first via email, and then I included the instructions again in the first Facebook post) about how to download GBoard – the Google Keyboards, for those who decided to use this program, and explained how to use ASR for the purposes of the study. GBoard is a keyboard app that supports voice dictation. It is freely available on the App Store and Google Play and so far, it has received positive reviews. Part of the students decided to use Siri or another voice dictation program that was already installed on their phones. Individual words were used for practice instead of sentences or phrases in order to minimize the program’s assumption out of context. The students practiced their pronunciation of words containing the vowel minimal pairs, for a period of two weeks, 20-30 minutes a day. They were instructed to start with one or two minimal pairs per session with the given vocabulary, and they were encouraged to add additional words containing the same sounds. They were encouraged to
add all the words (or as many as they can) in the last three days. I also explained that they should focus on the vowels and try to produce the correct vowel. The written output of the ASR program served as an indicator of their success. If ASR transcriptions were inaccurate, that was considered as feedback to the learners. The participants practiced individually, at a convenient time and place and occasionally posted updates about their progress and experience in the Facebook private group. Unfortunately, learning logs were not obtained because most of the students did not agree on taking this step because it would be too time-consuming for them, considering that they spent time practicing and posting in the Facebook group. The two-week practicing period is self-reported.

Phase 3 was similar to Phase 1. All the participants (EXP and NNSC) completed the Post-test by recording the same list of words containing vowel minimal pairs as on the Pre-test, and sent their recordings to the researcher through email. The learners from the EXP also provided their ASR written output, that is, the words which ASR wrote when they were producing the words at the end of the two-week practice, and sent the list to the researcher through email.

Finally, in Phase 4, the listeners transcribed the non-native speakers’ recordings from the Pre-test and the Post-test. The listeners were recruited through personal connections, contacted through email and were instructed to write down the words as they heard them, in normal English spelling. Each of them was assigned a separate Google document where the names of the recordings were written and an example was provided of what the transcription should look like. The recordings were given in a mixed order. In addition, the NSC provided a record of the ASR written transcription of their speech. The four phases are illustrated in Figure 3 for a better visual overview of the procedure.
Figure 3. The phases of the procedure

**IRB approval**

An Institutional Review Board (IRB) exempt study form was submitted to the Office for Responsible Research at Iowa State University and was approved on February 20th. Additional modification was submitted after receiving the Small Research Grant from the English Department for reimbursing $10 to the participants for their time, and was approved on March 18th.

**Data Analysis**

To answer the first and second research questions, about how learners improve their pronunciation after using the program, and how accurately the ASR feedback identify learner performance characteristics or errors, quantitative methods were used, including descriptive statistics.
In order to evaluate the usefulness of ASR for learners' proficiency gains, the Pre-test and Post-test speech samples of both EXP and NNSC were transcribed by native Listeners. The native listeners, were given instructions to transcribe the words exactly as they hear them in a normal English spelling. The recordings were given in a mixed order, so that the raters were not aware which recordings were from the post-test or pre-test, or which ones belonged to a control or experimental group. I manually calculated the accurately transcribed vowels, summarized them and divided with the total number of instances of that sound and turned into percentages. The focus on the analysis was on the vowel sounds, each sound containing three instances per speaker and 33 instances total per each vowel sound for the Experimental Group (EXP), and 30 instances for the control group (NNSC). The total tokens analyzed was 2640 for the pre-test and 2640 for the post-test for EXP and 2400 pre-test and 2400 post-test NSC. Errors with consonants sounds were not considered in this analysis. For example, if the listener transcribed ‘leaf’ and the target word was ‘leave’ the instance was considered as correct because the vowel sound was correctly identified. Finally, the overall accuracy of all the learners was calculated for the Pre-test and Post-test. I also identified the mean accuracy and the standard deviation (SD). The results from the Pre-test and the Post-test were compared in order to measure whether the learners had improved after the practice period or not.

ASR accuracy for transcribing the correct vowel sounds was measured by making two comparisons. First, the NNS written output from the ASR (provided by EXP) and the NS written output in ASR (provided by NSC) were compared. The total number of tokens analyzed was 264 (EXP) and 240 (NSC). I manually calculated the correctly identified vowels sounds, summarized them, divided them with the total number of instances and
turned into percentages; also calculating the mean score and SD. The same coding procedure was followed in coding the rest of the data. The purpose of this comparison was to identify and compare how the program worked for NS and NNS subjects. Second, a comparison was made between the ASR written output of NNS (provided by EXP at the end of the training, the same as above) and the native listeners’ transcription of their speech on the post-test. A Pearson correlation was used to measure the relationship between the mean scores of ASR recognition and the native listeners' recognition of non-native speech.

To answer the third research question, the last part of the analysis included qualitative analysis of students’ Facebook posts on their comments about their experience and attitudes towards ASR’s usefulness. A general inductive approach was used (Thomas, 2003) in which the raw data was framed around important key themes. Throughout the posts, several key themes arose. The students’ posts were copied and pasted in a Google document for the analysis and all the personal information were removed. To analyze the data, the posts were coded for emerging themes, concepts and beliefs, and hence conclusions were drawn. For example, the parts of the posts where the learners expressed positive attitude towards the system, parts where they expressed skepticism, parts where they discussed the usefulness of the system, and so on, were coded in different colors and comments to facilitate the analysis.
CHAPTER 4. RESULTS

This chapter discusses the results of the study. To answer the three research questions, there are three sections, each addressing a research question. The first section presents the results in regards to proficiency gains; that is, how much the students have or have not improved after the treatment period. The second section discusses the accuracy of ASR in transcribing speech and providing feedback. Finally, the third section provides a description of the learners' attitudes regarding their experience with using ASR.

Vowel pronunciation improvement

To answer the first research question, that is, to what extent learners improve their pronunciation after using ASR, 10 native listeners transcribed the pre-test and post-test recordings. Results (Table 3) showed that while EXP improved their overall accuracy score, the NNSC did not show any improvement.

Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>NNSC</td>
<td>62.08</td>
<td>22.77</td>
</tr>
<tr>
<td>EXP</td>
<td>58.33</td>
<td>25.33</td>
</tr>
</tbody>
</table>
Figure 4. Listeners’ accuracy scores for NNSC and EXP (M)

The average improvement for EXP of 6.71% showed that the learners slightly improved by using the program. As shown in Figure 4, the learners who used ASR (EXP) showed different results from the learners who did not have any pronunciation training but continued to attend their regular EFL classes (NNSC). While NNSC showed only .16% difference between pre-test and post-test, the EXP improved 6.71% difference between the pre-test and post-test which suggests that the training with ASR may have resulted with certain progress. While the overall (global) improvement is important, the purpose of the study was to investigate individual vowel pronunciation improvement. The individual percentage of accurate vowels, as recognized by Native Listeners, were summarized in Tables 4 and 5.
The results from the NNSC (Table 4) show that there is no notable difference between the pre-test and the post-test. In fact, in a few cases, such as /æ/, /u/, /ɑ/ and /ʌ/ the post-test results are slightly lower than the pre-test, 5.67%; 4%; 6.33% and 1.34%, accordingly. As for the sounds /i/, /ɪ/ and /ʊ/ the participants show slight improvement (1.66%; 1.67% and 4%), while a bit higher improvement (8.67%) was noted regarding the sound /ɛ/. Overall, the NNSC’s scores from the pre-test and post-test point out that no notable improvements were found for individual vowel sounds, with the exception of /ɛ/ (8.67%).

Findings regarding the EXP demonstrate that the learners improved their pronunciation of each vowel to some extent (Table 5). However, certain improvements were
higher than others. With a difference of 13.34%, /u/ appears to be the sound where the participants improved the most, followed by /æ/ with 12.73%, and /ʌ/ with 11.15%. Other improvements were found in the sounds /i/ and /ɑ/, both with 6.67% improvement. Finally, the difference between the pre-test and post-test results of the sounds /i/ and /ɛ/ were close to zero, 0.6% and 0.3%, respectively.

Even though the practice period was short, certain improvements are clearly visible from the comparison of the pre-test and post-test results, especially regarding the sounds /u/, /æ/, and /ʌ/. The notable improvement in the sound /u/ is important, because the English sound /u/ partially overlaps with the Macedonian /u/, indicating that it could cause difficulties in creating a new category for this sound for the learners. The learners showed the lowest scores regarding the pronunciation of the sound /æ/ (both on pre-test and post-test) which appears to be the most problematic sound for Macedonian EFL learners. The learners were the most accurate with the sounds /ɛ/, (93.94% pre-test and 94.24% post-test) which is almost completely overlapping with the Macedonian /e/; and the sound /ʌ/, whose pronunciation even further improved (11.15%) after the practice period.

**ASR Accuracy for vowel pronunciation practice**

In order to answer research question 2 of how accurately the ASR recognize native and non-native speech and how it relates to human raters, two comparisons were made: a) the learners’ written output in the ASR program was compared to human judgments (native listeners who transcribed their speech); and b) ASR written output of native and non-native speakers was compared. The data was manually coded for accuracy and the instances were marked as correct or incorrect based on the vowels. Percentages for the accuracy of the program were calculated for each vowel (by dividing the total accurate items by the total
number of tokens and turning them into percentages); as well as calculated the mean accuracy \((M)\) and the standard deviation \((SD)\).

Table 6.

*Mean Recognition Scores*

<table>
<thead>
<tr>
<th>Speakers’ L1</th>
<th>ASR Recognition Scores</th>
<th>Native Listeners’ Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Macedonian (N=11)</td>
<td>56.75</td>
<td>18.34</td>
</tr>
<tr>
<td>English (N=10)</td>
<td>78.75</td>
<td>12.13</td>
</tr>
</tbody>
</table>

Findings show that the overall ASR recognition of Macedonian speech was 56.75% while the native listeners’ transcription recognized 65.04% of their speech (Table 8). The difference between the ASR’s ability to transcribe or recognize the non-native speech and the native listeners’ ability to recognize the non-native speech was 8.29%. The recognition difference was not very large. Regarding the overall scores of recognition, ASR and human listeners showed similar performance, which may suggest that, even though there isn’t a complete match, ASR scores are closely related to human listeners’ scores.

On the other hand, a considerable difference was found between the ASR recognition scores for native and non-native speakers. The native speakers who provided their ASR written output belonged to the NSC, while the non-native speakers who provided their output belonged to the EXP and their output was collected after the training period. The recognition score for EXP was 56.75% while for NSC, it was 78.75%. The overall recognition score was not as high as expected for native English speakers, possible due to ASR’s issue with recognition of certain sounds or words; or ASR transcription of the more commonly used words.
Besides the overall scores, this study explored individual vowel recognition. To answer part a) of the second research question of how is ASR related to human raters, the accurate recognition scores for each vowel was calculated. A summary of the recognition scores (ASR and Native listeners) for each vowel is illustrated in Table 9.

Table 7.

<table>
<thead>
<tr>
<th>Lexical items</th>
<th>Leave</th>
<th>Live</th>
<th>Pan</th>
<th>Pen</th>
<th>Luke</th>
<th>Look</th>
<th>Cop</th>
<th>Cup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feel</td>
<td>Fill</td>
<td>Laugh</td>
<td>Left</td>
<td>Pool</td>
<td>Pull</td>
<td>Dock</td>
<td>Duck</td>
</tr>
<tr>
<td></td>
<td>Sheep</td>
<td>ship</td>
<td>ed</td>
<td>Bed</td>
<td>Fool</td>
<td>Full</td>
<td>shot</td>
<td>shut</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vowels</th>
<th>ASR M (SD)</th>
<th>Native Listeners M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>75.76 (2.61)</td>
<td>55.76 (15.99)</td>
</tr>
<tr>
<td>æ</td>
<td>45.45 (3.17)</td>
<td>29.09 (15.35)</td>
</tr>
<tr>
<td>e</td>
<td>33.33 (2.34)</td>
<td>29.09 (15.35)</td>
</tr>
<tr>
<td>u</td>
<td>78.79 (2.34)</td>
<td>94.24 (15.99)</td>
</tr>
<tr>
<td>o</td>
<td>66.67 (2.58)</td>
<td>71.52 (15.99)</td>
</tr>
<tr>
<td>a</td>
<td>36.36 (2.27)</td>
<td>66.67 (18.39)</td>
</tr>
<tr>
<td>ʌ</td>
<td>42.42 (2.61)</td>
<td>32.73 (21.67)</td>
</tr>
<tr>
<td>ɪ</td>
<td>75.76 (2.27)</td>
<td>96.06 (9.87)</td>
</tr>
</tbody>
</table>

Findings showed that ASR recognition was closer to native listeners for some of the vowels, while it showed lower recognition for others. The vowels /i/, /æ/, /u/ were similarly recognized from ASR and Native listeners with a difference of 1.52%, 4.24% and 4.85%, respectively. The difference in recognition for /ɪ/ was 10.31% and /ɑ/ it was 9.69%, which also falls into the acceptable difference defined in the hypothesis of this study (10-12% difference). However, native listeners had a higher recognition than ASR for /ʊ/ (30.31%) and /ɛ/ (15.45%). The situation was similar with /ʌ/, where native listeners recognized 96.06% of NNS, while ASR recognized only 75.76%. The number of tokens analyzed by native listeners is higher because the speech samples were rated by 10 native listeners.

Next, a Pearson correlation was computed to assess the relationship between these two variables, that is, ASR written output and native listeners transcription. The Pearson
correlation \( r = 0.84 \) showed that these two variables are closely related. Even though the native listeners showed greater recognition the ratings followed a similar trend (Figure 5) which suggests that ASR’s non-recognition may be identifying actual mispronunciations.

\[ \text{Figure 5. ASR and Native listeners’ recognition scores} \]

In addition, to answer the second part of RQ2 of how accurate ASR is in recognizing native and non-native speech, native speakers’ written output was collected (group NSC) and compared to non-native speakers’ written output (group EXP). ASR showed higher recognition percentages for native speakers compared to ASR recognition of non-native speech. Ideally, the program is expected to transcribe native speech with nearly 100% accuracy, or as defined in the initial hypothesis, the expected acceptable native speech recognition should be above 90%. However, in this study, ASR showed 96.67% recognition for native speech only for the sounds /i/ and /æ/. The program showed the lowest recognition with the sounds /ɔ/ and /ʌ/ (63.33%), while it ranged from 76.67% to 83.33% for /i/, /æ/, /ɛ/ and /u/.
Table 8.

**ASR Recognition scores for NNS and NS**

<table>
<thead>
<tr>
<th>Lexical items</th>
<th>Leave</th>
<th>Live</th>
<th>Pan</th>
<th>Pen</th>
<th>Luke</th>
<th>Look</th>
<th>Cop</th>
<th>Cup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feel</td>
<td>Fill</td>
<td>Laughed</td>
<td>Left</td>
<td>Pool</td>
<td>Pull</td>
<td>Dock</td>
<td>Duck</td>
</tr>
<tr>
<td>Sheephapship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vowels</th>
<th>i</th>
<th>i</th>
<th>æ</th>
<th>e</th>
<th>u</th>
<th>o</th>
<th>α</th>
<th>θ</th>
</tr>
</thead>
</table>

| NNS % (n=264) | 75.76 | 45.45 | 33.33 | 78.79 | 66.67 | 36.36 | 42.42 | 75.76 |
| M (SD)       | (2.61) | (2.97) | (3.17) | (2.34) | (2.58) | (2.27) | (2.61) | (2.27) |

| NS % (n=240) | 96.67 | 76.67 | 83.33 | 80.00 | 77.00 | 63.33 | 63.33 | 96.67 |
| M (SD)       | (4.88) | (3.99) | (4.32) | (4.42) | (4.12) | (3.72) | (3.98) | (4.88) |

Figure 6 depicts the results for each individual vowel as recognized by native listeners (non-native speech) and ASR (native and non-native speech). The figure visually shows the differences in recognition scores for individual vowels.

![Figure 6. Recognition scores for individual vowels](image)

This section made two comparisons: a) ASR written output of non-native speech vs. native listeners’ transcriptions of the non-native speech; and b) ASR written output of native speech vs. ASR written output of non-native speech. The results for the former showed high
correlation \((r=0.84)\) of ASR and native listeners in recognizing most of the sounds. The results from the latter showed that ASR has a higher recognition of native speech as opposed to non-native speech. Nonetheless, the program’s overall recognition of native speech was not as high as expected \((78.75\%)\). While the findings did not show high recognition for all of the native speech sounds, there was \(96.67\%\) for two of the sounds. Overall, we can conclude that the program has room for improvement in recognition of certain sounds even for native speakers, but, the results were positive in the comparison of ASR with native listeners’ recognition.

**Learners attitudes and beliefs towards ASR**

To address research question 3, regarding the learners’ attitudes towards using ASR for pronunciation training, the EXP participants discussed their experiences with ASR in a small closed Facebook group. A general inductive approach was used to qualitatively analyze the data.

Results showed that the participants’ attitudes towards vowel pronunciation practice with the use of ASR was generally positive. The Facebook group discussions showed that most of the learners were initially skeptical about the use of the program because they thought ASR systems are not highly reliable. Seven students reported that their expectations from the program were not very high. Nonetheless, they became more interested as they started using it. After receiving the instructions and explanations on how to use ASR with the goal to practice their vowels, most of the students reported that they knew ‘what to focus on’.

Their comments, in the Facebook group showed satisfaction when they became successful with ‘getting the program to write the word’ they were practicing. One of the learners said: ‘I kept trying to get the program to write the word *pan*. I said it more than 20 times.. and yes I succeeded and I was so proud of myself. I realized that I was pronouncing
the words *pen* and *pan* the same. I didn’t even know there is a difference between them.’

Some of the students reported that they became aware of the difference between the minimal pairs, commenting that before the training they were not aware that any difference existed. Throughout their posts, several of them reported that they looked up the words on YouTube to ‘hear the difference.’ However, they did not report if they managed to hear the difference between the vowel sounds. One learner commented: ‘…now at least I know they are different! I looked up the words on Youtube and online dictionaries…’. Even though this study did not examine perception improvement (because ASR did not provide any input to the participants), this may suggest that having the learners' awareness raised may lead towards the use of other sources and paying closer attention when listening to native speech.

Another relevant theme that emerged throughout the comments was that the participants in this study appreciated having the tool available at any time and being able to privately practice, without ‘being judged’. Receiving pronunciation feedback in front of their peers would make them feel demotivated to use English in class, and therefore they found using ASR to be their preferred method for receiving feedback. They also agreed that this tool is practical because they all have ‘some kind of ASR’ already installed on their phones. A few of them reported that they started using ASR as a means of communication with their friends, instead of texting (typing) because they thought it was ‘fun.’

Even though the general attitude towards the program was positive, occasional frustration was also reported. Four students reported that the program would write a completely different word because that made them feel unsure whether they were mispronouncing the word or the program had problems recognizing it. Two other students commented that they thought the program was more useful when the words were used in
sentences. For example, one of them said: ‘I couldn’t write the word pool right, it was either pull or oh or call or nothing on the screen. When I said a sentence like I jumped in the pool then it was fine.’ An interesting comment by another learner revealed that in a few occasions, they were not sure if they had a problem with the consonant or the vowels sounds when the program would ‘mishear’ their pronunciation.

Overall, the learners’ attitudes towards the use of ASR for vowel pronunciation practice showed that learners appreciated having a tool to help them with pronunciation practice even though it was not always reliable. Considering that none of the participant had previous pronunciation training, the use of this tool may have helped to raise their awareness of vowel pronunciations and lead towards an autonomous search for additional tools.
CHAPTER 5. DISCUSSION AND CONCLUSION

This study investigated the usefulness of ASR for vowel pronunciation practice by exploring learners’ vowels pronunciation improvement after practicing with the program, the relationship between ASR recognition and rater’s recognition of non-native speech, the accuracy of the software for transcribing native and non-native speech, and learners’ attitudes towards using ASR. This chapter summarizes and discusses the findings and makes connections to previous research in this field. In addition, implications, limitations, and directions for future research are discussed. Finally, conclusions are drawn at the end of this chapter.

Discussion

The initial hypothesis of this study stated that the learners would show improvement after the treatment period; ASR would show an acceptable level of accuracy (approximately 10-12% difference as compared to human judgment and above 90% recognition of native speech, therefore, its feedback will be useful, and the learners would have positive experience and attitudes towards using the program. To evaluate the overall usefulness of mobile-assisted ASR programs, all these aspects of usefulness need to be considered. Past literature, examining one or two of these aspects, lead to the creation of this hypothesis. To test the hypothesis, several calculations and analysis were done. In order to see whether it has been supported, the results regarding each research questions will be discussed.

The results of the first part of this study show that the experimental group (EXP) improved their overall accuracy of vowel pronunciation as compared to the control group (NNSC) that did not show any improvement. Even though the overall improvement of pronunciation was not particularly high, possibly due to the short two-week training, some
progress was evident. These results go in line with Mroz (2018) who also found that the learners improved their pronunciation after using a mobile-assisted tool. Other studies that support similar findings regarding segmental improvement are Wang and Young (2015), Cucciariini et al., (2009) and Neri et al., (2008). These findings suggest that learners who use ASR for feedback might improve their pronunciation.

While other studies mostly compared the overall (global) improvement of the learners’ speech before and after the treatment period, this study looked at the improvement of each of the eight selected vowels in four minimal pair contrasts /i/, /ɪ/; /æ/, /ɛ/; /u/, /ʊ/; and /ɑ/, /ʌ/. While the overall scores are an appropriate measure for assessing the global improvement of the learners’ pronunciation, looking at each individual vowel can provide a better insight into the learners’ actual improvement. The EXP showed varying improvement for most of the vowels. Liakin et al. (2015) also explored individual sound acquisition (for the French sound /y/) and found that the learners improved after using mobile-assisted ASR. The learners in the current study improved the most problematic sound for Macedonian EFL learners /æ/ from 16.36% accurate pronunciation in the pre-test to 29.09% in the post-test, followed by improvements with the sounds /u/, /ʌ/, /ɪ/ and /ɑ/.

These gains may be attributed to a number of factors that may include the production of pronunciation output. Swain’s Output Hypothesis (Swain, 1985) suggested that language production was important in three main ways: triggering noticing; receiving feedback and conscious thinking about the speech production. The language production in the current study using ASR for providing feedback may have led to noticing of different vowel contrasts because learners received individual feedback on their speech, and they may have started to consciously think about their pronunciation of different vowels. Other factors that
are involved in the pronunciation improvement gains, as discussed by Liakin et al. (2015) based on research by Dabaghi (2010), Dekeyser (1993), Rosa & Leov (2004), Christison (1999) and Chun & Plass (1996) may be the effect of an explicit focus on the target form (practicing of vowels), immediate feedback (from the ASR system) and multiple opportunities for learning. Nonetheless, while longer-term improvement would depend on many other factors, including students’ aptitude, attitude, personal motivation and willingness to improve their pronunciation, it is crucial to start by raising the awareness of the existence of unfamiliar sounds.

Another important aspect for evaluation of the ASR systems which was investigated in this study is the level of accuracy in the process of turning the speech into text. In terms of accuracy, the standard that the ASR systems should be measured against is that of human judgments (Levis, 2007). The results regarding the first part of the second research question, that is, how accurate the program is as compared to human judgment, showed that ASR and human judgment had similar recognition with an acceptable difference of 8.29%. These findings suggest that human raters marked similar mispronunciation errors as ASR. In other words, it is very likely that ASR’s inaccuracy of transcribing the non-native speech might have been pointing out the learners’ mispronunciation because it was very similar to the human judgment of inaccurate vowels. Even though overall scores are a suitable measure because the ultimate goal is to improve overall pronunciation quality, it is also important to provide in-depth analysis when assessing ASR systems (Neri et al., 2006a). In this study, I looked at each individual vowel and how they were recognized by ASR and the human listeners. ASR showed an acceptable level of recognition for most of the vowels as compared to human judgment. It showed lower recognition level for the sound /ʊ/ (30% difference
between ASR and listeners), followed by /ʌ/ and /ɛ/ (20% and 15% lower than listeners, respectively). Overall, ASR and human judgments seemed to follow a similar trend (Figure 5) and appeared to be relatively closely related ($r=0.84$, $p<0.01$). These findings are supported by O’Brien et al. (2018) who argue that early studies that primarily dealt with pronunciation assessment using ASR demonstrated relatively high correlations between human judgments and machine scores (referring to studies by Cucchiarini et al., 2000a, 2000b, 2002; Neumeyer, Franco, Digalakis, & Weintraub, 2000; Witt & Young, 2000). These results were also supported by Cucchiarini et al., (2009) who pointed out that the scoring accuracy of their ASR-based system was relatively good, as assessed by a human expert, and Coniam (1998) who also found that ASR goes consistently in line with human raters and it is most reliable on a word level. The findings of this study point out that ASR recognition scores may be a good indicator of the learners’ mispronunciations, as they aligned closely with the native listeners’ judgments for most of the vowels.

As for the second part of RQ2, when evaluating the accuracy of ASR, it is also important to test the level of recognition for native speakers. Commercial ASR systems were originally designed for native speakers (Levis, 2007) and they may show a lower recognition level for non-native speakers. The findings in this study showed 78.75% recognition for native speakers and 56.75% for non-native speakers. The percent for native speakers’ recognition is surprisingly lower than expected, as part of the existing research showed a higher level (up to 95%) of ASR recognition for native speakers (Levis, 2007). The lower recognition for native speakers might suggest that these programs have certain issues with recognizing words written in isolation, or problems with recognition of certain sounds. Nonetheless, the lower ASR recognition for non-native speakers (56.75%) may be due to the
learners’ mispronunciations. Other studies found from 65 to 72% accuracy for non-native speakers (Ashwell & Elam, 2017; Coniam, 1999; Derwing et al., 2000). In this study, the lower recognition of non-native speech is considered to be the feedback that learners would receive from the program, which is also supported by the alignment of the ASR recognition scores with native listeners.

As seen in Table 10 and Figure 5, for native speech recognition ASR was the least reliable for the sounds /ʊ/ and /ɑ/ (63%), while for the rest of the sounds it ranged from 77 to 97%. As discussed above, the sound /ʊ/ appeared to have the biggest difference between ASR recognition for non-native speech and native listeners. We can also note that when it comes to the sound /ɛ/, ASR recognition of native and non-native speech was almost the same (79% and 80%) while it was recognized by 94% from the native listeners. This sound appeared to be the easiest sound for Macedonian learners, based on the comparison between the Macedonian and English vowel systems, yet ASR’s ability to transcribe it correctly (for both native and non-native speech) seems to be limited around 80%. Interestingly enough, the sound /ɛ/ was more commonly ‘misheard’ by ASR and replaced with the sound /ɪ/, for both native and non-native speakers. For example, the word ‘pen’ was occasionally incorrectly transcribed as ‘been’ or ‘pin,’ the word ‘left’ as ‘lift’ and the word ‘bed’ as ‘bit’ and in one instance as ‘that.’ One possible reason for the inaccuracy of the ASR system for these sounds, especially for native speakers, might be due to overlapping in the pronunciation of the vowels /ɛ/ and /ɪ/ in some of the regional varieties in the USA (Clopper, Pisoni, & De Jong, 2005).

As for the non-native speakers, mispronunciation of the initial consonant might be one reason that led to a higher inaccuracy of the system. For example, the plosives /p, t, k/
are unaspirated in the Macedonian language (Kirkova-Naskova, 2011) and that may have
caused misinterpretation of the whole word. The non-native speakers also reported that they
noticed ASR occasionally had problems with recognition of words with initial /p/ and /f/
sounds and final /l/ sound. More specifically, the words 'pool', 'fool' and 'full' were
transcribed as ‘call' or ‘phone' or 'pool,' interchangeable in a few instances. Another reason
for the inaccuracy with these sounds may also be connected with the final /l/ sound. As
reported by the learners in the Facebook discussion, ASR did not have any problems when
transcribing these words in a sentence, such as ‘The glass is full.' This may be one of the
most important shortcomings of the ASR programs in this study. The program sometimes
may rely on context, and for that reason, it may have some issues in recognizing individual
words. It is also possible that the program is more accurate for words, but it may use things
other than pronunciation for sentences. In any case, if the program recognizes the words
more easily from the context, using full sentences for practice may be pointless, unless the
words are the same parts of speech, for example: ‘I told her a joke and she left.’ Vs. ‘I told
her a joke and she laughed.’ In addition, the lower recognition of the words containing the
sound /ɑ/ (cop and dock) might have appeared because the words are less commonly used as
opposed to the /ʌ/ words (cup and duck) according to their frequency in the Corpus of
Contemporary American English (COCA).

Overall, despite the notable limitations in accuracy, especially the lower ASR
recognition of native speech in this study for particular vowels, the high level of alignment
between native listeners’ and ASR’s level of recognition, looks promising for this kind of
pronunciation practice. Even if the ASR system is not 100% accurate, it appears to be similar
to the native listeners’ perception which may indicate that vowel pronunciation practice using
this type of mobile-assisted commercial ASR (Gboard, Siri or voice recognition on smartphones) could provide similar feedback as native listeners.

Last but not least, referring back to the third research question, the learners’ attitudes are also an important factor in evaluating the usefulness of ASR. The Facebook group discussion posts showed that the EXP group’s attitudes towards ASR were generally positive. The learners’ initial skepticism in using ASR for pronunciation practice was very likely due to their expectations of lower recognition of non-native speech. The learners were not highly satisfied with previous attempts to turn speech into text and hence hesitated in the use of this kind of system for improving their pronunciation. Nonetheless, the ‘flaw’ of ASR with low non-native speech recognition was used as a feedback of mispronunciation in this study. When the learners received the explanation and instructions on how to use this program, they decided to give ASR a chance to help them improve.

Throughout the practice, learners reported general satisfaction with using ASR. Three of them said that they found the activity ‘useful’ and four of them reported that they are satisfied with the feedback they received from this program. Similar findings by Mroz (2018) also showed that learners appreciated ASR feedback because for them it provided a visual representation of their words. She also found that the program seemed to raise awareness of one’s own intelligibility and also outward awareness in what makes speech comprehensible. A similar discussion developed between the EXP members. Some commented about not knowing the difference between most of the minimal pairs before this training, mostly focusing on /æ/ - /e/ and /u/-/ʊ/; and one of them commented that they ‘did not care about pronunciation before now.’ Wallace (2015) also argued that students “have credited this activity with opening their eyes and their ears” (p. 183) to the types of difficulties that people
could have to understand them. It is very likely for students who had minimal or close to zero
pronunciation training, to have their pronunciation awareness raised by the use of ASR.
Other studies have also reported positive attitudes towards ASR (Chiu, Liou, & Yeh, 2007;
Cucciariini et al., 2009; Mroz, 2018; McCrocklin, 2016) which supports the findings in this
study.

Some of the learners also reported looking for other outside resources with the goal to
hear the pronunciation of the vowels that were problematic for them. This may be an
indicator that the use of ASR promoted more autonomous learning. Having their awareness
of vowel pronunciation raised by the use of ASR allowed them to observe and correct their
own errors (Ahn & Lee, 2016) as well as increased motivation and willingness to look for
additional resources to improve their perception of the words. Previous studies showed that
ASR is useful for fostering autonomy (McCrocklin, 2016), also highlighting the potential use
of commercial ASR used on mobile phones in EFL context (Ahn & Lee, 2016; Liakin et al.
2015; Mroz, 2018). ASR can be useful for both ESL and EFL context. Nonetheless, while in
ESL context the learners have contact with native speakers and are exposed to native
language in their surroundings, ASR can be especially useful for EFL learners where
exposure to native language is limited (McCrocklin, 2016).

Even though occasional frustration was reported with the use of ASR, the students
referred more to benefits than drawbacks from its use. The private environment that allowed
them to have multiple attempts to practice receiving immediate feedback as well as the
practicality of having the program with them anywhere (and for free), kept the learners’
positive attitude towards using it. In a lack of pronunciation resources that are freely
available in small countries such as Macedonia, lack of instructional time for pronunciation,
and lack of exposure to native speech, ASR may be a useful tool to raise students’ awareness of mispronunciation of vowels.

To sum up, going back to the initial hypothesis in this study, the learners from the experimental group (EXP) showed improvement after the treatment period; they showed positive attitudes towards the use of ASR for vowel pronunciation practice; and ASR showed (overall) an acceptable level of accuracy. ASR and native listeners had a relatively strong relationship and followed a similar trend. The only part of the analysis that was not strongly in favor of ASR was the low recognition of native speech (78.75%). Besides the global lower recognition, the program showed flaws with particular sounds. Overall, the initial hypothesis was partially supported by this study (ASR's native speech recognition lower than expected being the part which makes the hypothesis not fully supported). Besides the lower recognition of native speech, the exploration of the other aspects of ASR usefulness was in favor of ASR. Practicing vowel pronunciation using a mobile-assisted ASR tool on smartphones may allow the students a number of possibilities for pronunciation practice that they typically do not have. With careful guidance and targeted practice, the use of ASR may lead to pronunciation awareness raising and possible pronunciation improvement. This does not suggest that ASR dictation practice should be a complete substitute for classroom instruction (McCrocklin, 2019). Rather, ASR can serve as a tool to free up classroom time and facilitate more autonomous practice at home for the learners. The teachers cannot just leave the learners to their own devices but instead, they need to provide clear goals and guidance to the learners to make this type of practice work for their classrooms. Therefore, this study supports the use of ASR for vowel pronunciation practice for EFL environment with careful guidance and observation from the instructor.
Limitations and directions for future research

It is important to acknowledge the limitations of this study before making wider generalizations. While the number of participants was limited, the number of tokens analyzed for the pre-test and post-test from native listeners was relatively high, having 4800 tokens for the NNSC (2400 pre-test and 2400 post-test) and 5280 (2640 pre-test and 2640 post-test) for the EXP group. Nonetheless, the analyzed tokens from the ASR output were much lower (240 NSC and 264 EXP). The reason for the difference between the number of tokens in these two analyses is that the number of tokens increased having 10 native listeners as raters, while only one set of ASR output was provided. For future studies, it would be advisable to obtain more data from ASR output to get more reliable results. Another limitation of this study is the use of the word ‘laugh’. The word was occasionally pronounced as the British over the American variety and was therefore transcribed as ‘lof’ by the native American listeners. For future research, I would recommend avoidance of words that have different pronunciation in different accents of English. Moreover, due to IRB recommendation, the learners were given freedom in their choice of the type of mobile-assisted ASR they used. For future studies, exploring only one type of ASR would give a clearer picture of the accuracy of that particular ASR program, as their recognition level may be different.

In addition, while the Facebook group, which was open for discussion for the students, was useful for following their progress throughout the process, taking a survey at the end of the training and a delayed survey would possibly provide a more structured and straightforward analysis. Another limitation of the study was the short training period of two weeks. For future studies, an extended training period which ideally would be integrated into EFL courses can provide more authentic environment where the teacher can play a central role in combing use of ASR with classroom instructions. In addition, the learners’ daily
practice in this study is self-reported as the learners felt it would be too time-consuming for them to write daily logs. For future research, daily logs and an extended number of words would help strengthen the claims. An additional measure of the ASR usefulness would be to measure the improvement of untrained words by limiting the practice to a list of words. Adding words that were not trained in both pre and post-test can show whether the learning transfers to new words. This may allow the researchers to see whether the improvement had generalized to other words that were not included in the training.

**Implications**

This research has implications for both teachers and researchers. First, this study shows that ASR can be useful for raising the learners' awareness about the contrast between minimal pairs. Bearing in mind that pronunciation resources, as well as teacher's time in the classroom, might be limited in an EFL setting, integration of ASR can help raise the learners' awareness and possibly motivate them to work autonomously on improving their pronunciation. The learners in this study showed improvements in all of the vowel sounds even after a short training period, which may indicate that integration of ASR training in the classroom throughout the school year, may be beneficial for the students. Most importantly, students had positive attitudes towards using ASR. They found this activity to be fun and engaging, and some of them reported that they have started using the speech-to-text feature for communication with their friends. Having a tool that is already available on their smartphones (which they use daily), could facilitate the process of practicing allowing them a safe environment where they will receive feedback but will avoid the (possibly) intimidating pronunciation correction in front of their peers. This may allow them to self-observe, to
analyze their weaknesses, and lead them to search for additional resources that can help improve their perception and hence their pronunciation more generally.

Nonetheless, it is also important to point out that the learners in this study showed skepticism at the beginning of the training and occasional frustration when they could not get the program to write the right word. With careful guidance from the teacher and a focused pronunciation training, the students can learn how to use the program for the right purpose.

Conclusions

This study explored the use of mobile-assisted ASR by investigating three aspects of its usefulness, that is, the degree to which ASR can promote improvement, the accuracy of the program as compared to human judgments and to native speakers’ recognition, and the learners’ attitudes towards using the program for vowel pronunciation practice. The findings indicate that the learners from the EXP gained overall (global) proficiency as well as improvements for each individual vowel. The findings also showed that ASR’s written output was closely related to human judgment with an acceptable difference between them (less than 10%). Nonetheless, the comparison of human judgment and ASR recognition was not uniformly similar for all vowels. Despite these limitations, the students appeared to have positive attitudes towards ASR and enjoyed the training.

While ASR still has room for improvement in terms of accuracy with native speech recognition, this type of training can serve as an awareness-raising tool that could be integrated into the L2 classrooms. Therefore, this study recommends careful guidance from the teachers, focused and structured practice using individual words and use of mobile-assisted ASR programs for more practical experience. This study was the first that explored
Macedonian EFL learners’ use of ASR, and it provides directions for future research for the use of ASR in EFL contexts where native input is limited.
REFERENCES


APPENDIX A. LIST OF MINIMAL PAIRS FOR ANALYSIS

/ɪ;/i/
Live/ Leave
Fill/ Feel
Ship/ Sheep

/ɛ;/æ/
Pen/ Pan
Left/ Laughed
Bed/ Bad

/ʌ;/ɑ/
Cup/Cop
Duck/Dock
Shut/Shot

/ʊ;/u/
Full/ Fool
Pull/ Pool
Look/ Luke
APPENDIX B. LINGUISTIC BACKGROUND SURVEY

1. Name/Pseudonym
2. Gender
3. How old were you when you started studying English?
4. How many years have you studied English?
5. Have you been to an English-speaking country (USA, UK, Canada etc)?
   a. No b. Yes _____ Length of stay
6. Which languages other than English have you studied?
   a. Italian b. French c. German d. Other _________
7. Have you ever had specific English Pronunciation training?
   a. No b. Yes, please explain __________________
8. What is your English proficiency? (Self-assessment)
APPENDIX C. IRB

The project referenced above has been declared exempt from most requirements of the human subject protections regulations as described in 45 CFR 46.104 or 21 CFR 56.104 because it meets the following federal requirements for exemption:

2018 - 2 (ii): Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) when any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation.

2018 - 3 (ii.B): Research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses or audiovisual recording when the subject prospectively agrees to the intervention and information collection and any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation. If research involves deception, it is prospectively authorized by the subject.

The determination of exemption means that:
• You do not need to submit an application for continuing review. Instead, you will receive a request for a brief status update every three years. The status update is intended to verify that the study is still ongoing.

• You must carry out the research as described in the IRB application. Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, nature or duration of behavioral interventions, use of deception, etc.), any change in privacy or confidentiality protections, modifications that result in the inclusion of participants from vulnerable populations, removing plans for informing participants about the study, any change that may increase the risk or discomfort to participants, and/or any change such that the revised procedures do not fall into one or more of the regulatory exemption categories. The purpose of review is to determine if the project still meets the federal criteria for exemption.

• All changes to key personnel must receive prior approval.

• Promptly inform the IRB of any addition of or change in federal funding for this study. Approval of the protocol referenced above applies only to funding sources that are specifically identified in the corresponding IRB application.

Detailed information about requirements for submitting modifications for exempt research can be found on our website. For modifications that require prior approval, an amendment to the most recent IRB application must be submitted in IRBManager. A determination of exemption or approval from the IRB must be granted before implementing the proposed changes.

Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

Additionally:

• All research involving human participants must be submitted for IRB review. Only the IRB or its designees may make the determination of exemption, even if you conduct a study in the future that is exactly like this study.

• Please inform the IRB if the Principal Investigator and/or Supervising Investigator end their role or involvement with the project with sufficient time to allow an alternate PI/Supervising Investigator to assume oversight responsibility. Projects must have an eligible PI to remain open.
• **Immediately inform the IRB** of (1) all serious and/or unexpected **adverse experiences** involving risks to subjects or others; and (2) any other **unanticipated problems involving risks** to subjects or others.

• **Approval from other entities may also be needed.** For example, access to data from private records (e.g., student, medical, or employment records, etc.) that are protected by FERPA, HIPAA or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **An IRB determination of exemption in no way implies or guarantees that permission from these other entities will be granted.**

• Your research study may be subject to **post-approval monitoring by Iowa State University’s Office for Responsible Research.** In some cases, it may also be subject to formal audit or inspection by federal agencies and study sponsors.

• **Upon completion of the project, transfer of IRB oversight to another IRB, or departure of the PI and/or Supervising Investigator, please initiate a Project Closure in IRBManager to officially close the project.** For information on instances when a study may be closed, please refer to the [IRB Study Closure Policy](#).

Please don’t hesitate to contact us if you have questions or concerns at 515-294-4566 or [IRB@iastate.edu](mailto:IRB@iastate.edu).