Guskaroska, A. (2019). The potential of ASR for facilitating vowel pronunciation practice for Macedonian learners. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 250-260). Ames, IA: Iowa State University.

PRESENTATION/POSTER

THE POTENTIAL OF ASR FOR FACILITATING VOWEL PRONUNCIATION PRACTICE FOR MACEDONIAN LEARNERS

Agata Guskaroska, Iowa State University

The purpose of this study is to examine Automated Speech Recognition (ASR) software and its potential for facilitating vowel pronunciation practice for Macedonian English as a Foreign Language (EFL) learners. A list of 12 sentences including minimal pairs of the contrasts /i/-/ɪ/, /æ/-/ɛ/, /u/-/ʊ/, and /ɑ/-/ʌ/ was recorded by 10 Macedonian learners, aged 18-19 and two American English native speakers in order to test the reliability of ASR. The speech samples were turned into text using ASR and the results of the written output were compared between native speakers and non-native speakers. Results demonstrated that the program was accurate in transcribing most of the vowel sounds for native speech. ASR written output was less accurate for non-native speech and was most likely indicating learners' mispronunciations of vowels by transcribing them inaccurately. The results suggest that ASR may be promising for individual vowel practice but future research may involve words in isolation to avoid the system's flaws in making assumptions based on context.

INTRODUCTION

The process of second language acquisition requires development of several aspects of the second language. One of the areas which is usually neglected by instructors, possibly due to lack of time, desire, or training, is pronunciation (Huensch, 2018). Learners often express a desire to work on their pronunciation (LeVelle & Levis, 2014; McCrocklin & Link, 2016), but pronunciation is a skill that requires feedback and is difficult to acquire autonomously (McCrocklin, 2016). In this digital era, researchers are exploring technology with the aim of finding appropriate tools that can assist L2 pronunciation improvement by providing feedback to learners (Levis & Suvorov, 2014; Wallace, 2016).

In that regard, several studies have explored the effectiveness and the potential of ASR (such as *Dragon NaturallySpeaking, Google web speech, and Siri*) and its ability to assist learners by providing feedback with the text-to-speech written output (Derwing, Munro, & Carbonaro, 2000; McCrocklin, 2016; Mroz, 2018). Levis & 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. Findings from previous studies (e.g., Coniam 1998; Derwing, Munro, & Carbonaro, 2000; Eskenazi, 1999) have mostly indicated that ASR was not fully developed to provide reliable feedback to the learners. Nonetheless, these researchers agreed that if ASR were to improve in the future, it could provide a wide range of possibilities for language learning.

Nonetheless, recent studies found generally positive results towards use of ASR for pronunciation practice (Liakin, Cardoso, & Liakina, 2014; McCrocklin, 2016; Mroz, 2018). These studies pointed towards establishment of learner autonomy and progress. ASR has tremendous potential

in applied linguistics and learners appreciate its use (Levis & Suvorov, 2014). It looks promising for pronunciation self-access work and can provide a safe environment for learners. While past research is mostly in favor of ASR, researchers pointed out that the software's accuracy needs further exploration. In that regard, more research is needed to examine whether ASR has improved throughout time. Therefore, this study will examine Macedonian learners' use of ASR as a way to test the system's accuracy.

Contrast between the Macedonian and the English vowel system

The phonetic system of the Macedonian standard language includes five vowels: /i/, /e/, /a/, /o/ and /u/. In English, there are arguably around 12 vowels and eight diphthongs (Dodd & Mills, 1996). In this paper we focus on the American accent variety. Because there are many more vowels in English than in Macedonian, almost every English vowel presents a potential pronunciation problem for Macedonian learners and may be classified as non-existent in the Macedonian language (Kirkova-Naskova, 2012). Even /ɛ/ which is acoustically closest, is pronounced differently depending on phonetic context and dialect region the Macedonian learner belongs to. Figure 1 depicts a comparison between Macedonian and English vowels diagrams.

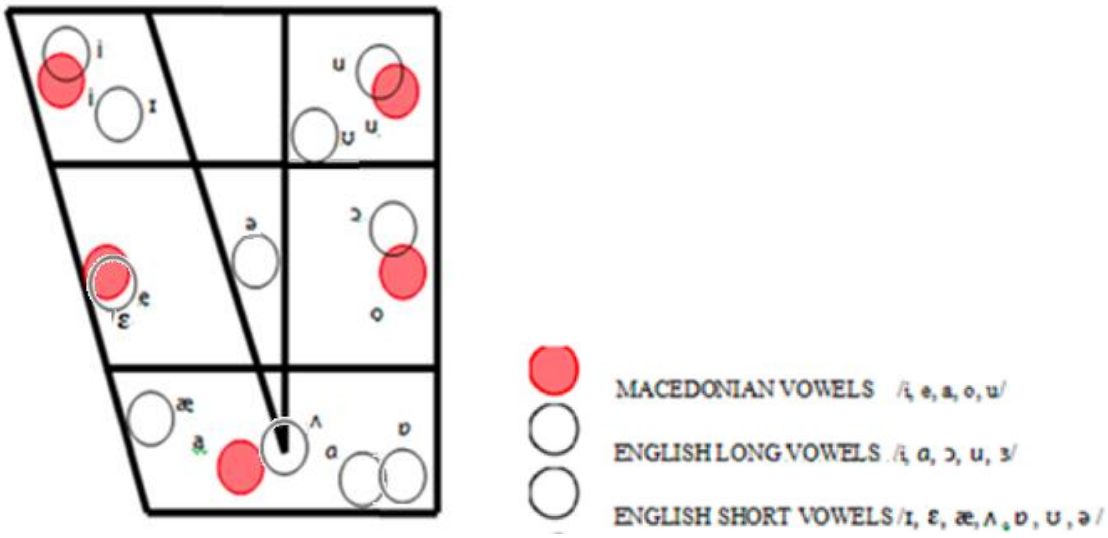

Figure 1. Macedonian and English vowel diagrams (adapted from Krikova-Naskova, 2012).

Based on the comparison above, the first selected vowel pair that might be problematic for Macedonian EFL learning is /i/-/ɪ/. This contrast is very frequent and difficult for these learners because, in Macedonian, there is only one sound which is somewhere in between these two sounds (Kirkova-Naskova, 2009). The minimal pair /æ/-/ɛ/ is also an important contrast because /æ/ does not exist in Macedonian while /e/ is the closest with the English /ɛ/. The /u/-/ʊ/ contrast is similar to /i/-/ɪ/ in terms of difficulty of perception by Macedonian learners. Even though the contrast between these two sounds is not frequent in English and has a low functional load (Munro & Derwing, 2006), Macedonian learners rarely hear the difference between these sounds. Finally, none of the sounds /ɑ/ and /ʌ/ exist in Macedonian and learners very often substitute them with the Macedonian sounds /o/ or /a/, respectively. Hence, this study includes the following vowel contrasts: /i/-/ɪ/, /æ/-/ɛ/, /u/-/ʊ/, and /ɑ/-/ʌ/.

The study

Inspired by the lack of research in this field, as well as lack of tools to assist the EFL/ESL classrooms, this study investigates the potential of an ASR tool, *Apple's Enhanced Dictation Feature*, for providing corrective feedback to Macedonian learners. *Apple's Enhanced Dictation* is available in OS X Mavericks v10.9 or later. This program is free, easily available, and user-friendly and for those reasons it was selected for this study. This study explores ASR's accuracy by comparing ASR's recognition of native and non-native speech. The exploration of the written output of NSs will show the accuracy for these speakers which can help in the exploration of ASR's potential to facilitate vowel pronunciation practice for Macedonian learners.

Research questions

The following questions guide this study:

1. How accurate is the ASR program Enhanced Dictation in recognizing and transcribing native English speakers' production of vowel contrasts?
2. How accurate is the ASR program in recognizing and transcribing Macedonian L2 learners' production of vowel contrasts?

METHODOLOGY

Participants

The participants who took part in this research were Macedonian EFL learners, aged 18-19, who provided speech samples of their English for evaluation. 10 Macedonian native speakers, seven male and three female (level B2, n=4; and C1, n=6) according to the Common European Framework of Reference for Languages, were recorded. None of the participants had lived for any period in an English-speaking country.

The control group included two male, native speakers (NS) of American English, graduate students familiar with pronunciation, aged 28-29. The NSs speech was recorded to provide a standard for comparison in order to evaluate the program and its ability to turn words into text.

Materials and procedure

The materials consisted of 12 sentences containing minimal pairs that were the same parts of speech (e.g. "The patient wanted to leave."; "The patient wanted to live."). See the Appendix for the minimal pairs. These minimal pairs were deliberately chosen to be the same parts of speech to avoid the program's assumptions of certain words based on their position in the sentence. Based on the comparison between the English vowel systems, the selected sounds were included (/i/-/ɪ/; /æ/-/ɛ/; /u/-/ʊ/; and /ɑ/-/ʌ/). All the vowel contrasts had three instances containing the vowel, for example, for the sound /i/, the words *leave, sleep* and *sheep* were chosen. The selected words consisted of simple vocabulary that is introduced at low levels of EFL classes and hence known to the students to avoid problems in pronunciation due to lack of knowledge of the word meaning. ASR used for turning the voice into speech for this study was *Apple's Enhanced Dictation*.

The participants were encouraged to first read the sentences once to themselves quietly and then to record themselves while they read the sentences aloud at a normal pace. Participants recorded their speech with a voice recording application on their phone (iPhone or Android) and sent the recordings via email. The speech samples were played to ASR, the speech was turned into text and saved as text in a Microsoft word document.

Data analysis

The words were manually evaluated for accuracy, then counted separately per vowel and total and turned into percentages. In order to evaluate the program's accuracy for recognizing native and non-native speech, the data were separately analyzed and summarized in tables. After that, a comparison of the results was made between NS and NNS's written output to identify differences and similarities.

FINDINGS

Recognition of native speakers of English

To answer the first research question, the ASR's written output of NSs speech was analyzed. During the analysis, the focus was only on the targeted minimal pair words, not the entire sentence they were embedded in. The sentences only served to provide context because the purpose of this study was to focus on vowel contrasts.

Table 1

Number and percent of accurate and inaccurate recognized lexical items by the ASR program (Native English speakers)

No. of participants	No. of lexical items per speaker	Total No. of lexical items	Accurate	Inaccurate
2	24	48	42 (87.5%)	6 (12.5%)

Overall, the program did not provide 100% accuracy when it comes to vowel recognizing and turning voice into speech for NSs, in the context used. The program examined showed 87.5% total accuracy, which is close to what several similar studies found. For instance, Derwing et al. (2000) found 90% accuracy and Ashwell and Elam (2017) found 89.4%. Even though the total accuracy of the program in this study is 87.5%, analyzing each sound recognition individually can provide us with a clearer picture of the tool's capabilities. Table 2 provides a closer look into each sound and identifies the sounds which the program failed to recognize.

Table 2

Number and percent of accurate and inaccurate recognized vowel contrasts by the ASR program (Native English speakers)

Lexical items	Leave Sleep Sheep	Live Slip Ship	Pan Laughed Man	Pen Left Men	Luke Wooed Boot	Look Would Book	Cop Dock Shot	Cup Duck Shut
Vowels	i	ɪ	æ	ɛ	u	ʊ	ɑ	ʌ
Accurate items								
No.	6	6	6	4	6	6	6	2
%	100%	100%	100%	66%	100%	100%	100%	33%
Inaccurate items								
No.				2				4
%				33%				66%
Total								
No.	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Interestingly, almost all the NSs' vowels were transcribed 100% correctly with the exception of the vowels /ɛ/ and /ʌ/. Regarding the vowel /ɛ/, the only lexical item that ASR did not recognize was the word *men*. The sentences used for this commonly mistaken vowel contrasts were: *I saw the man with the yellow coat* and *I saw the men with the yellow coat*. ASR failed to recognize the plural form of this word in all the instances, which made the recognition of /ɛ/ 66% accurate. If it is not just a challenging pair, perhaps the system relies on context to assist in word recognition. In other words, the system may suppose the singular form of the word and thus transcribes the word as *man* in both sentences. Future research could explore the accuracy of the program by isolating the words and not providing any context. On the other hand, the program transcribed all the other /æ/-/ɛ/ words correctly, thus proved accurate in this study by 100% for recognizing /æ/ and 66% for /ɛ/ sound.

Another vowel that was unrecognized was /ʌ/. Only 33% of the words containing the vowel /ʌ/ were recognized and transcribed correctly. The issue with the recognition of these sentences may be an indicator of the program's assumption based on context. For example, the sentence *I sat on the duck* was contrasted to the sentence *I sat on the dock*. One possible explanation is that the word *dock* may likely appear more often in this type of context and hence the program may have transcribed the word incorrectly merely making an assumption based on frequency. Regardless of the reasons, these findings show that the ASR program did not appear to be highly reliable for the sound /ʌ/ used in this context. In this study the overall accuracy of the system's recognition of NS vowels appears to be high, nonetheless, it might be important to consider vocabulary and context selection in order to avoid the system's possible limitations.

Recognition of non-native L1 Macedonian ESL speech

To answer the second research question, I calculated the number and percentage of recognized vowels in the system's written output of the targeted minimal pairs of NNS. The overall score of

accuracy for NNS was 71%, as seen in Table 3. These findings align with Derwing et al. (2000) study where they found that the software was 71-73% accurate for nonnative speech for Cantonese and Spanish L1 learners, while Ashwell and Elam (2017) found 65.7% for Japanese and a few Chinese speakers.

Table 3

Number and percent of accurate and inaccurate recognized lexical items by ASR (Macedonian ESL learners)

No. of participants	No. of lexical items per speaker	Total No. of lexical items	Accurate	Inaccurate
10	24	240	170 (71%)	70 (29%)

Nonetheless, the overall results present the systems' accuracy in general, and do not give a clear picture about each targeted vowel. Table 4 summarizes the findings for each sound separately to get a better overview of the situation.

Table 4

Number and percent of accurate and inaccurate recognized vowel contrasts by ASR (Macedonian EFL learners)

Lexical items	Leave Sleep Sheep	Live Slip ship	Pan Laughed Man	Pen Left Men	Luke Wooed boot	Look Would book	Cop Dock shot	Cup Duck shut
Vowels	i	ɪ	æ	ɛ	u	ʊ	ɑ	ʌ
Accurate items								
No.	22	22	17	20	16	29	29	15
%	73%	73%	57%	67%	53%	97%	97%	50%
Inaccurate items								
No.	8	8	13	10	14	1	1	15
%	27%	27%	43%	33%	47%	3%	3%	50%
Total								
No.	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

When looking at individual sounds produced by the L2 learners, we can note that no sound was recognized with 100% accuracy, although the sounds /ʊ/ and /ɑ/ are close, both at 97% recognition. The lowest percentage of accuracy was with /ʌ/ with 50% accuracy. However, when comparing to the NSs, the system was not considered reliable regarding the sounds /ʌ/ by transcribing only 33% of NS words correctly. This may suggest that the overall system struggled to recognize this sound. Other sounds with low recognition were /u/ with 53% and /æ/ with 57%. Both /i/ and /ɪ/ showed 73% accuracy. As discussed earlier, this pair was expected to be difficult for Macedonian learners. However, the pair is also very frequent which might have resulted in better pronunciation then other sounds. On the other hand, when it comes to the sounds /æ/ and /u/, ASR demonstrated 100%

accuracy for native speech and only 43% and 47% accuracy, respectively. These findings suggest that these Macedonian learners may have issues with distinguishing the production of most of the vowel contrasts, considering that ASR transcribed NS accurately.

DISCUSSION

To be useful for vowel pronunciation practice for L2 learners, ASR should first recognize native speech with high accuracy. The overall score of recognition was lower than expected with 87.5% accuracy (see Table 1). The non-native speech was transcribed less accurately (71% as shown in Table 3). Is this an indicator that the program cannot transcribe a non-native speech, or is it an indicator that the program gives good feedback to the learners because it writes what it 'hears'? This percentage might be interpreted as the general ability of ASR to indicate intelligibility and, as Wallace (2016) points out, to suggest the words which were unclear. Even though previous studies criticized the ability of ASR to recognize non-native speech (Coniam, 1999; Derwing et al., 2000), more recently ASR tools have been improving and several recent studies are in favor of the program for L2 pronunciation practice (Liakin et al., 2014; McCrocklin, 2016; Mroz, 2018; Wallace, 2016).

The overall results may align with previous studies, however, the overview of individual sounds shows that almost all the vowels were transcribed 100% correctly for NS with the exception of the vowels /ɛ/ and /ʌ/. Regardless of whether the selected pairs might have been challenging or the program might have 'assumed' words out of context, with the exception of these two vowels, ASR showed 100% accuracy for the rest of the vowels for NSs. These findings may be indicators that in future studies, vowel pronunciation practice should be tested by using individual isolated words, instead of sentences, to eliminate the possibility of the influence of context.

On the other hand, ASR did not show the same level of accuracy for identifying individual vowels for Macedonian learners. Was the program identifying vowel mispronunciations? While it cannot be claimed that these errors were due to mispronunciation, as it may be due to other reasons, closer analysis of the output showed that many of the errors seem closely connected to problems with mispronunciation. Kirkova-Naskova (2010) also points out that the most challenging minimal pairs for Macedonian learners is /æ/-/e/, also identifying /u/-/ʊ/,/ɪ/-/i/ and /ʌ/-/a/ as common foreign markers in Macedonian-accented speech. In this study, ASR seemed to be identifying specific vowels that were likely mispronounced by these speakers and present the most common foreign markers of their speech. Hence, it could be argued that the program appeared to be providing feedback to the learners' mispronunciations and might be considered useful for vowel pronunciation practice for Macedonian learners. In order to confirm this hypothesis, future studies should include NSs' judgments of the non-native speech.

Previous studies on ASR pointed out that it can be beneficial to students in various ways, such as creating a safe environment for self-practice, saving time, self-monitoring (Wallace, 2016), fostering learner autonomy, supplementing course work (McCrocklin, 2015), and raising students' awareness (Mroz, 2018). Mroz (2018) found that learners are mostly satisfied with their ASR experience, emphasizing that the written output was a good feedback for them as it provided visual representation of their words. All these benefits make ASR an interesting field that needs further exploration.

This exploratory study for Macedonian learners of English for vowel pronunciation practice showed that, besides exploring the overall accuracy scores, examining the way individual sounds are turned into text can also be valuable and should also be explored when evaluating ASR's accuracy. The findings suggest that ASR was most likely indicating learners' mispronunciations of vowels by transcribing the words inaccurately. Hence, this study may provide evidence that ASR has promising potential for L2 learners vowel pronunciation practice and should be explored further in the future.

CONCLUSION

This study explored the accuracy of an ASR tool, *Apple's Enhanced Dictation*, and its possibility to provide corrective feedback for vowel pronunciation practice in an EFL context. Even though the enhanced dictation feature is limited to Macintosh users, the results suggest that ASR may have great potential for providing corrective feedback to EFL learners for a select set of vowel contrast. Even though the overall accuracy score for NSs was not as high as desired, the program was accurate in this study for recognition of individual vowel sounds for American native speech. The only sound for which the program demonstrated flaws was the sound /ʌ/ (only 33% correct). In terms of Macedonian EFL speech, the ASR written output was less accurate and it was most likely indicating learners' mispronunciation of vowels by transcribing the words inaccurately.

For future studies, words containing the target vowel sounds can be used in isolation to avoid possible influence of the sentence context when the program turns speech into text. Furthermore, to confirm the usefulness of the program, future studies may include native human raters in order to make a comparison between the program's feedback and human judgment. Finally, ASR may be recommended for individual vowel practice for Macedonian EFL classroom use, but further research is required to confirm these findings.

ACKNOWLEDGMENTS

I would like to thank Dr. John Levis for his guidance and support. In addition, I would like to thank Dr. Shannon McCrocklin, Dr. Lara Wallace and Erin Todey, for their valuable feedback and observations on this topic.

ABOUT THE AUTHOR

Agata Guskaroska is a Fulbright scholar and a MA student in TESL/Applied Linguistics at Iowa State University. She taught ESL courses for seven years and American literature at FON University. Her major interests include CALL, pronunciation, and SLA.
Email: agatag@iastate.edu

REFERENCES

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

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27,* 49-64.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular recognition software work with ESL speech? *TESOL Quarterly, 34,* 592-603. https://doi.org/10.2307/3587748

Dodd, S., & Mills, J. (1996). *Phonetics and phonology.* CORE, University of Exeter Press. core.ac.uk/display/19510000.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, *2*(2), 62–76. http://dx.doi.org/10125/25043

Huensch, A. (2018). Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Language Teaching Research*. 1–20. DOI: 10.1177/1362168818767182

Kirkova-Naskova, A. (2012). Interlanguage phonology: comparison between the English and the Macedonian vowel systems. *Annual Symposium of the Faculty of Philology 'Blaze Koneski'* pp. 141-152. Skopje: 'University St. Cyril and Methodius'.

Kirkova-Naskova, A. (2010). Native Speaker Perceptions of Accented Speech: The English Pronunciation of Macedonian EFL Learners. *Research in Language, 8,* 1-21.

Kirkova-Naskova, A. (2009). *Markers of Foreign Accent in Macedonian-accented English*. Unpublished MA Thesis. Skopje: Faculty of Philology.

Levelle, K., & Levis, J. (2014). Understanding the impact of social factors on L2 pronunciation: Insights from learners. In J. Levis, & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 97-118). Boston: DeGruyter.

Levis, J., & Suvorov, R. (2014). Automated speech recognition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Retrieved from *http://onlinelibrary.wiley.com/store/10.1002/9781405198431.wbeal0066/asset/wbeal0066.pdf?v=1&t=htq1z7hp&s=139a3d9f48261a7218270113d3833da39a187e74*

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1-25.

McCrocklin, S. (2015). Automatic speech recognition: Making it work for your pronunciation class. In J. Levis, R. Mohammed, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 126-133). Ames, IA: Iowa State University.

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

McCrocklin, S., & Link, S. (2016). Accent, Identity, and a Fear of Loss? ESL Students' Perspectives. *Canadian Modern Language Review, 72*(1), 122-148.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals,* 1-21. https://doi.org/10.1111/flan.12348

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*(4), 520-531. https://doi.org/10.1016/j.system.2006.09.004

Wallace, L. (2016). Using Google web speech as a springboard for identifying personal pronunciation problems. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 180-186). Ames, IA: Iowa State University

APPENDIX.

Minimal pair sentences

/u/-/ʊ/

1. **Look/Luke** Look! There's a rabbit over there. Luke! There's a rabbit over there.
2. **Full/Fool** What's the meaning of the word 'full'? What's the meaning of the word 'fool'?
3. **Boot/Book** I lost my boot. I lost my book.

/i/-/ɪ/

4. **Leave/Live** The patient wanted to leave. The patient wanted to live.
5. **Sleep/Slip** Did you sleep on the ice? Did you slip on the ice?
6. **Sheep/Ship** Where's my sheep?/ Where's my ship?

/æ/-/e/

7. **Men/Man** I saw the man with the yellow coat. I saw the men with the yellow coat
8. **Pen/Pan** Can you please give me the pen? Can you please give me the pan?
9. **Left/Laughed** I told her a joke and she left/ I told her a joke and she laughed.

/ʌ/-/ɑ/

10. **Cup/Cop** I don't like that cup. I don't like that cop.
11. **Duck/Dock** He sat on the duck. He sat on the dock.
12. **Shut/Shot** The door was shut. The door was shot.