TImproving Students' Pronunciation Using Auto Speech Recognition Software to the Tenth Grade Language Major Students at SMAN 1 Lawang

Muhammad Subkhan
English Language Teaching Study Program
Postgraduate School, Universitas Islam Malang, Indonesia
Email: han_muhammad.3@gmail.com

Abstract

Clear pronunciation is important in order to communicate well in English. Unfortunately, the ten grader students of language major at SMAN 1 Lawang find difficulties to pronounce English words in classroom activity. This research is classroom action research because the aim is to figure out a solution to solve the pronunciation problem experienced by ten grader students of language major at SMAN 1 Lawang with the help of proper teaching media. The technology of auto speech recognition is the most suitable program for pronunciation improvements. This research is using one of the auto speech recognition which namely "Speech Text & Text Speech in All Languages" developed by SoftWiz. This research applied audio feedback activity with several adjustments. The adjustments were the use of auto speech recognition program that operated using android mobile phone and the use of google voice to replace the tape recorder voice. This research involved 31 ten grade students of language major. The result was a consistent improvement of students' pronunciation in every session which was proven by the consistent class's average score improvement, therefore, it proves that the strategy was arguably successful to be a solution of the students' pronunciation problem.

Keywords: Pronunciation, Audio Feedback, Auto Speech Recognition

INTRODUCTION

The Language is the main and the important part of communication. The communication process will not go well without better language skills. Hrehova (2010) stated that language represents a significant foundation of communication. In order to communicate with everyone all over the world, international language existence is a must. English is considered as a language that is used by people all around the world and becomes the most frequently used language internationally. Therefore, English is chosen as an international language. McKay (2002) stated that English as an international language is used by multilingual societies in international communication.

Unfortunately, there were problems that emerged related to mastering English for Indonesian young learner specifically in senior high school level. Generally, the problems were curriculum adjustments that was not support the need of students, the numbers of English materials and times that was not well-balanced, the teacher and students ratio in a classroom which were contributed to the lacked of development in English classroom activity.

In order to figure out the solution for those problems, the researcher was conducted a study in the only state high school in Lawang district namely SMA N 1 Lawang. Based on the early observation, the researcher found out that the students of SMAN 1 Lawang in language major found difficulties to pronounce word correctly. The problem emerged because of students' lack of confidence. The students' low confidence was caused by the dominance of skillful students who often answer most of the question and grab most of the opportunities given by the teacher to show their performance. Therefore, the researcher proposed a strategy to deal with the problems which was using the help of auto speech recognition. The strategy was chosen to provide a chance to self-learning that supported by recent technology. Considering the current development of the technology that grows greatly, teachers cannot stick with old media such as flashcard, board games, etc. The teacher needs to get to

know the high-tech media such as speech recognition that can integrate to the teaching and learning process in the class.

Speech recognition is a computer based system that allows humans to communicate with the computer and the computer will respond to the words spoken by the human. The response given by the computer may vary such as typed the word uttered or do the users' command based on the word uttered depends on the users' need. The speech recognition technology works in orders. The first is recording the words pronounced by humans using the microphone. Then the computer analyzes the phoneme uttered, before matches the word uttered to the database and decides the closest word that is likely produced by the human. Speech recognition technology is claimed to have great impacts in language learning activity. With suitable adaptation with the need of language learning process, the speech recognition provides great opportunity for students to improve their skills both inside and outside the class. Students have a chance to increase their fluency as well as their confidence by using speech recognition technology.

Studies have done previously to figure out how auto speech recognition technology supported the English classroom activities. The research entitled "Improving English Pronunciation via Automatic Speech Recognition Technology" is proving that the use of speech to text technology has a great impact on students' speaking development. The research has been done by Meihui Li, Meiting Han, Zejia Chen, Yiling Mo, Xiujuan Chen, and Xiaobin Liu (2017).

Based on the research, it is difficult for Chinese to be able to speak in English correctly as there are dozens of different dialects in China that have different ways of pronunciation as English. In order to help Chinese students to improve their English pronunciation, the researcher decided to use ASR (Automatic Speech Recognition). The use of ASR in second language education, especially English, in China is common and it is considered as an effective way to improve Chinese ability to speak in English. The research was trying to explore the effectiveness of the use of ASR in improvement of English and the students' attitude towards the ASR. It was taking twenty-nine freshmen from School of Foreign Studies of South China Normal University as the research subjects who already learn English for around 9 to 12 years but not quite aware of the lesson.

The research needed 4 weeks and used both pre-test and post-test. Both tests were using the similar text. At the beginning of the research, the students were asked to read the text provided using ASR in which the technology was able to transform their speech into text. The reading session needs 20 minutes before the students' allowed continuing to read. The students were required to take a picture of the result of their speech that emerged on their phone screen. Based on the picture, students were asked to figure out their mistakes and try to redo until they were able to pronounce correctly. The tests were taken every week to record the students' progress.

The researchers then conclude that the use of ASR is able to help students to improve their English and has a significant impact on them. The use of ASR will help students to learn the correct English pronunciation by themselves. It also helps teachers in China to provide better equipment in English speaking class. Unfortunately, the ASR technology is not widely used in China just yet due to several limitations. One of the limitations is that most schools in China are forbidding the students' to bring the phone to school.

There is another research about the use of ASR in the same year as the research that has been done by Chinese researchers. In 2017, research about the use of ASR has been done towards Iraqi students. The research entitled "The Effect of Automatic Speech Recognition EyeSpeak Software on Iraqi Students' English Pronunciation: A Pilot Study" by Sidgi (2017). He claims that for recent classroom activity for foreign language, technology is playing a critical role in the development of the students' ability due to the limited time to practice the languages. ASR is one of the technologies that is suitable for foreign language classrooms as well as English. It is stated in his research that ASR now becomes an important tool in teaching speaking or pronunciation and gives massive help for teachers so that teachers can spot students' problem in pronouncing the words. His research is focused on investigating the effectiveness of the use of ASR in order to improve pronunciation of Iraqi students in the Department of English Language at Al-Turath University College.

The researcher claims that the use of ASR already differs the class from the traditional one in which in order to get better pronunciation, teachers will correct the students' mistakes directly. This may make students uncomfortable and lose interest in learning. Meanwhile, by applying ASR in the classroom, it will make students get private correction and feedback provided by the software. The research, as the previous research done in China, needs 4 weeks with one meeting every week for around 45 minutes. As the research is a pilot project, it only takes 10 students that are chosen randomly from the first year student of English Department with the age range from sixteen to twenty one years old.

Before getting experimented by the researcher, the student will follow a pretest session in order to understand their level of English, then they will be briefed by the researcher about how to use the software. The result of this research shows that ASR has a significant contribution towards students' development of English pronunciation. The students are highly improved on the use of /p/, /v/, /ʃ/, a slight improvement in /ʒ/, /tʃ/, /dʒ/ and less improvement in /ŋ/.

As a conclusion, the researcher points out that the research objective is to help Iraqi students to improve their pronunciation in English although there are differences in phonetic system between Arabic and English. The

use of ASR software is proven effective in improving students' ability in English; therefore the researcher is recommending the use of ASR especially for EFL learners.

Those previous studies supported this study in several ways. Firstly, both of the studies were using free auto speech technology that proven able to increased students' ability to pronounce word properly refers to native speakers, in this case was google default voice. The benefit of using auto speech recognition was that every student will able to get the application. Secondly, the students both of the previous research were applied in the classroom activities that were conducted during regular meeting. That proved the effectivity of the use of auto speech recognition application to support English classroom activity. Lastly, the result of both researches showed that the use of auto speech recognition able to improved students' pronunciation. The similarities between those previous researches and this research were the use of free auto speech recognition program although the name of the programs was different and the use of the program was during regular English classroom activity.

In this research, the researcher used speech recognition software named "Speech Text & Text Speech in All Languages" that is published and developed by SoftWiz developer. The "Speech Text & Text Speech in All Languages" application chosen based on its advantages that differ from other applications. "Speech Text & Text Speech in All Languages" is a free-download media that is already downloaded by more than ten thousands users and it can be downloaded by every android user. The applications rated for 3+ out of 5 by the users which mean that the application is rated above average. The application is easy to use with simple guidance. The best part of the application is that it offers a reader machine using google default voice. It means, student will be able to listen the correct and like-native pronunciation so that the student knows how to pronounce the word correctly.

METHOD

The research is considered as a classroom action research because this research intended to help students and teacher to formulate an effective teaching and learning strategy to solve the students' pronunciation problem. Kemmis& McTaggart (1998) stated that there are steps to do in classroom action research in order to complete the cycle, and it should not depended on one cycle only, which means there must be another cycle that needed to be done to improve the first cycle.

The researcher started the research with a preliminary research in the classroom using the auto speech recognition program directly. Based on the preliminary research, the researcher figured out several misspelled words. The researcher than prepared the steps to do for the research. This research applied several steps of classroom action research which consists of four steps in every cycle; planning the strategy, acting the strategy, observing the result, and evaluating the process

The researcher has done research planning in several steps, which were downloaded the application on google playstore and make sure the application can work well either for recording or play the default word pronunciation provided by the application. Next step was preparing the text that will be used in the classroom. The text was a recount text that was taken randomly from the internet that has been approved by the English teacher. The website posts the text was https://fungsi.co.id/recount-text/.

The third step of planning was counting the number words in the text. The chosen text consists of 124 words. Then the researcher did the last step to complete the research planning which was creating scoring rubric. The calculation of the score was as follows: correct pronunciation gets a score as much as 1, false pronunciation gets 0 score. The numbers of correct pronunciation divided by max score then multiply by 100. Once the research plan was ready, then it will be implemented in the classroom activity. The researcher makes an agreement with the English teacher of SMAN 1 Lawang that the research would get four sessions.

The sessions were conducted in following steps. First, the students are asked to open their application that has already downloaded previously, and take a look at the chosen text that is shared in the whatsapp group. The students are asked to read the text without recording it. Second, the researcher will ask ten students to record their reading and screen capture the result of the recording. The students must send the screen captured result to the whatsapp group by mentioning their student number.

Third, the result of the recording will be evaluated by the researcher. Based on the evaluation, the students are asked to type the words that they pronounce incorrectly, then click the play button so that the students will hear the correct pronunciation according to the application's native speaker. Once the students understand how to pronounce it, the students are asked to redo the activities. The result of the second attempt on recording the audio will also need to be sent to the whatsapp group and will be compared by the researcher. It also becomes the data to be observed later.

The research was claimed success once met the some criteria, which were; the application runs smoothly in students phone, the students able to follow the instruction very well, the students have score improvements consistently on four attempts, the students reach minimum score 75 of English.

RESULTS

The research's result was divided into two cycles. The first cycle was held in January 2020. All ten grader students registered as a student of language major in SMA N 1 Lawang were attending the class. Although the students already knew how to use the application, the researcher explained the way to use the application again to make sure the students understood the instruction clearly. The researcher explained the details of the applications, the icons, and the functions. The researcher also gave the students chances to ask questions about the applications. There were no questions asked. The researcher then sent a recount text that was already approved by the English teacher of the class to the whatsapps group that was created earlier so that every student would have it. Once every student has it, they ask to read the text carefully.

The second step of the cycle was recording students reading with the speech to text application. The students who already finished reading were asked to record their reading. The process of reading was in the class and the student read and recorded it together in that class. They recorded it using their own phone. There were three students that had problems with their application in which the application did not record their reading, those students then asked to use their friend's phone. Another student had a problem with her phone in which the application was not available for the type of her phone, and she was asked to borrow her friend's phone as well.

It needed around 30 minutes to finish the recording for every student. The students were then asked to screen captured the result of the recording and sent it to the group. The students were also asked to evaluate the result of their own recordings and learned the mistakes they have made. The students were allowed to play the audio of the misspelled words. After sometimes taking a look and evaluating their recordings, the students were asked to do one more recording. The students who had finished evaluating their recording were allowed to start the recording process. The process was the same with the first recording. The students asked to record the reading with the phone they have, screen captured the result, and sent it to the whatsapp group. The session was over once all of the students sent the result of their recording in the form of a screen capture of the recording. The researcher collected the image to evaluate it.Based on the evaluation done by the researcher, the score of the recording both in the first and second sessions are as shown on the table below:

Table 1The First Cycle Result

Student's Number	Session I score	Session II score	Difference
1	81	83	+2
2	88	80	-8
3	100	94	-6
4	32	84	+52
5	73	77	+4
6	94	91	-3
7	85	80	-5
8	94	97	+3
9	92	83	-9
10	96	92	-4
11	93	95	+2
12	64	78	+14
13	90	83	-7
14	87	90	+3
15	72	62	-10
16	96	98	+2
17	66	91	+25
18	90	90	0
19	97	70	-27
20	94	81	-13
21	94	85	-9
22	96	92	-4
23	57	83	+26
24	63	80	+17
25	75	81	+6
26	64	71	+7
27	73	86	+13
28	90	92	+2
29	74	70	-4
30	91	75	-16
31	90	92	+2
Average	82	84	+2

Based on the table above, some of the students got good results on the first session. Twenty students were able to score higher than eighty, while the rest of the students got lower than eighty. One student who reached a score higher than eighty, was able to get a perfect score. On the other hand, one student of those whose score was lower than eighty, was able to get the lowest score as thirty two. The average score of the first session was eighty two.

The second session started once the students finished re-read the text, examined their mistakes, and listened to the audio of the correct pronunciation of the words. Some students were able to reach a higher score than their previous score. Some others got lower scores from the first session. There were seventeen students who increased their scores, six of them greatly increased their score by more than ten points difference. The others increased the score by less than ten points difference. On the other hand, fourteen students decreased their score. Four of them decreased their score by more than ten points, while the others decreased by less than ten points. One student from thirty one students got an equal score from both sessions. The average score was also increased by two points in the second session. The average score was eighty four.

During the activity that has been done in the classroom, the researcher also conducted an observation. The researcher observed the activities by walking around the class and listened to some of the students' reading randomly. Based on the observation, the researcher found out several problems that should be fixed for the next circle. First problem was the environment. Once the students started their recording at the same moment, the class became noisy. The sounds were overlapped. Some of the recordings were not good because it recorded other students reading who sat near another student. This condition happened to the students who were speaking in low volume and got disturbed by students who were speaking louder.

The second problem emerged was the phone that was used by the students. Some of the students' phones could not record clearly enough although the words were correctly pronounced. The written words of the recording were showing other words that were not related. For instance, some of the recordings showed a sentence as "I miss you" that was supposed to be "I missed it". Some of the phones were not able to record the recording directly. Those phones were lagged and in the end unable to record the sentences.

The third problem that occurred in both the first and second session was the students' reading speed. Some of the students were reading the text fast that made their pronunciation become unclear, so that the application did not catch the pronunciation correctly. The researcher assumed that those students wanted to finish the recording as soon as possible so that they could do another activity in the class. Lastly, the researcher found difficulties controlling the class because the recording process of all of the students was done at once. The researcher could not give the proper feedback, since the researcher was unable to listen to the students reading carefully during the sessions.

Based on the data gathered in the form of text produced by students' recording and also the result of researcher observation, the next circle was modified to manage the first circle deficiencies. The modifications in the second circle were; students recorded their readings face to face with the researcher and found new room so that limited any other sound around the students that might interrupted the process of recording, used only researcher's phone to ran the application or recorded students' reading, and made sure the speed of reading was proper. Those modifications were reported to the English teacher and the teacher approved it

The second circle then underwent a week later. The students who had previously joined the first meeting were complete and none of them was absent. The researcher explained the modifications of how the meeting should be done to the students and told the students about the first meeting evaluation. The students were asked to follow the new instructions which were: come to the another classroom to conduct a recording in front of the researcher one by one, did the two sessions directly without going back to the class in between two sessions, using the researcher's phone to record their reading, decreasing their reading speed and paying more attention on the words pronunciation. The students who were not yet called got an assignment from the English teacher.

The second meeting had similarities with the first meeting in which students' were asked to read the text before recording it, every student had two sessions of recording, the result of recording was screen captured, saved, and evaluated later by the researcher.

Here was the result of second meeting:

Table 2The Second Cycle Result

Student's Number	Session I score	Session II score	Difference
1	94	78	-16
2	90	84	-6
3	97	98	+1
4	89	97	+8
5	90	75	-15
6	94	94	0
7	94	91	-3
8	93	92	-1
9	75	92	+17
10	90	98	+8
11	88	97	+9
12	81	73	-8
13	93	94	+1
14	99	85	-14
15	76	67	-9
16	98	100	+2
17	98	81	-17
18	90	98	+8
19	79	97	+18
20	58	100	+42
21	81	94	+13
22	89	99	+10
23	92	76	-16
24	90	80	-10
25	80	100	+20
26	89	65	-24
27	91	87	-4
28	92	98	+6
29	87	96	+9
30	85	94	+9
31	91	96	+5
Average	88	90	+2

Based on the table, the students in the second meeting were able to improve their reading as much as six points average compared to the first meeting from as same as two sessions from both meetings. The modifications of the strategy or research implementation successfully increased the quality of the recording so that the students were more focused and the result was better.

There were twenty seven students who reached a score more than eighty. It was a huge impact compared to the first session of the first meeting. The difference was seven students. As a consequence, the number of students that were unable to get a score more than eighty was greatly decreased to only four students. Unfortunately, there was no student who was able to get a perfect score. The lowest score was improved from thirty seven to fifty eight. The average score for the first session in the second meeting was eighty eight.

The students were once again asked to re-read the text before they were allowed doing the second session. It took around five minutes for the students to read the text again before completing the second sessions. Once the students were ready, the second sessions started.

There were seventeen students who were able to increase their score; three of the students were able to get the perfect score. There were six students who increased their score more than ten points difference, while the rest were increased lower than ten points difference. On the other hand, there were thirteen students who failed to keep their score and got decreased. Unfortunately, six students decreased their score more than ten points, while the others decreased lower than ten points. The good news was, the students' lowest score was sixty five which showed improvement from the other sessions.

The second cycle was gaining some successes, but there were also new problems that emerged. The first one was time consuming. The second circle needed more than two hours given by the teacher. It took approximately four hours to finish the circle. The time consuming here was because the meeting was done in face to face between the students and the researcher. As a consequence, the recording session was unable to start and ended at the same moment for every student. Moreover, the reading speed was limited therefore students were not allowed to finish it in a rush.

Second problem was the battery level of the researcher's phone. During the process of recording, the battery was getting low so that the researcher needed to find the phone replacement. One of the students was

offering his phone and the researcher approved the phone. The recording did not go well, later on the researcher figured out that the phone was just repaired. There was a problem with the microphone that was why the recording did not run well.

The researcher tried to find another phone replacement that did not have any problem. Fortunately, the researcher found one phone that worked well with the application as well as the recording. The phone was borrowed for the few students who did not yet do the recording. Although there were problems with the second cycle, the result was considered as success for the research. Considering the time limitation of the research and the result of the second circle, the third cycle is unnecessary.

Based on the two cycles that have been done by the researcher here is the chart of the students' achievements:

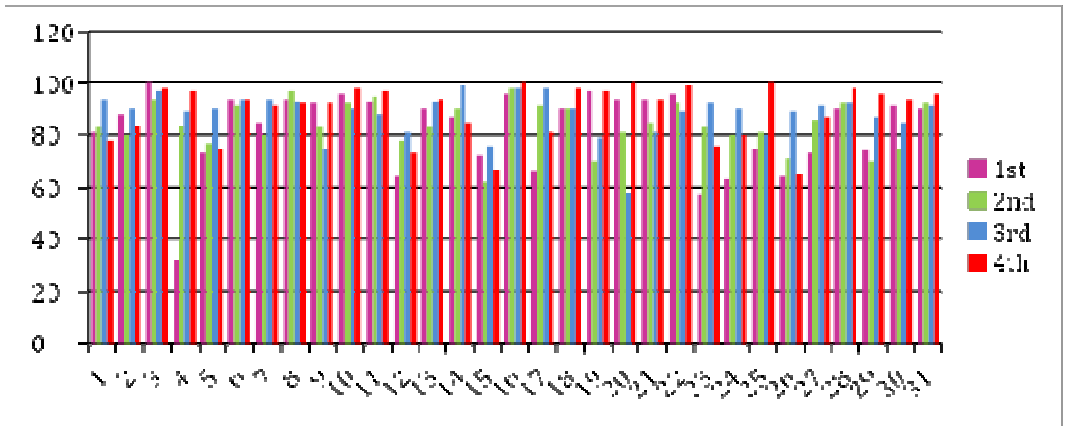

Figure 1 Students' Performances

The data shows that the number of the students who were experiencing the consistent improvements is seventeen students. The students are student number 3,4,9,10,11,13,16,18,19,20,21,22,25,28,29,30, and 31. The others were experiencing inconsistent performance. There are fourteen students who were inconsistent. They are the students number 1,2,3,6,7,8,12,14,17,23,24,26, and 27. The average score of all students' recording attempts are as follows:

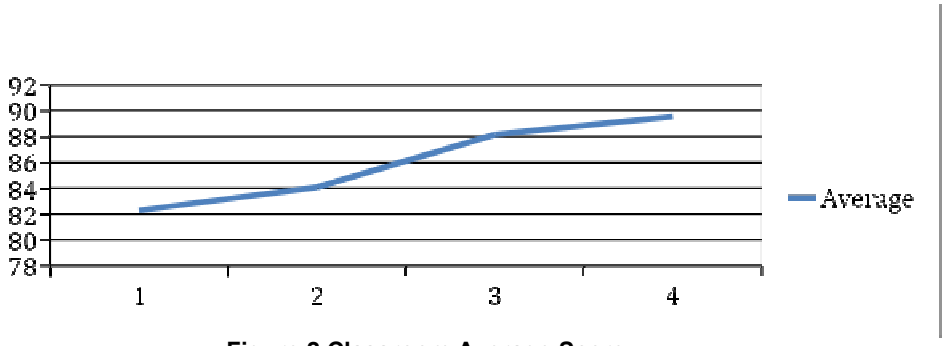

Figure 2 Classroom Average Score

Based on the table above, it is clear that the class average score is consistently increasing every time. The class average score in the first attempt of recording is eighty two. It increases to eighty four in the second attempt of recording. Greatly increase in the third attempt as much as eighty eight, and it reaches ninety.

The score improvement is very well done by the tenth grader students of SMAN 1 Lawang language major and it proves that the strategy is arguably successful to use in the classroom activities, particularly in speaking skills development.

DISCUSSION

Based on the result of both cycles, that was disadvantages as well as advantages applying the auto speech recognition in the classroom activities. Based on the result of the research and the observation in the class, there were several benefits that the teachers and the English classroom will gain from applying auto recognition speech technology using "Speech Text & Text Speech in All Languages" application. First, the use of technology in the classroom was proven to be helpful. The classroom that uses technology in the process of teaching and learning will mostly be able to improve or increase the result of the process. It is because both the students and the teachers are excited about it which will drive the mood of learning. Another reason, for a classroom that is not using recent technology in the process of teaching and learning will find something new and will increase the awareness of the students about the topic they learned.

Second, Speech Text & Text Speech in All Languages application in particular, is a part of recent technology which is easy to use and it is applicable in the English classroom activity in SMA N 1 Lawang and in any other schools that have similarities with SMA N 1 Lawang. It is because the application did not require high tech android, or the newest version of android. It requires the android that was released in 2012.

Third, the strategy of using Speech Text & Text Speech in All Languages applications is supporting self-learning methods. The application already provides its users with default google voice that will be able to pronounce the words like natives therefore, the users, in this case the students of SMA N 1 Lawang will have no limitation of learning pronunciation.

The fourth strength or benefit that is taken from this strategy is to keep the scoring objective. There is a clear scoring rubric to score practical lessons in English, which is speaking skills. By converting the speech into text, the teacher will be able to count down the number of mistakes that have been done by the students and the students will know their mistakes so that they will not make the same mistakes in the future. The teacher will not judge the students' performance based on their appearance, but it will focus on the speaking process.

On the other hand, there were some disadvantages of this research. The first weakness is, the application used to do the learning strategy is demanding internet access. The users of the technology must have an active internet connection on their phone. The internet is used to support some important features of the applications such as google default voice and automatic language translator. The application is still able to work without internet connection but with limited performance that possibly lead to misinterpretation of the reading which will convert into different words that differ from the original words.

The second weakness is that the application needs a quiet place with less sound around so that the application will be able to focus on detecting the users' voice and the third weakness is weakness is the reading speed. Using technology of automatic speech recognition is different from any other technology in terms of its proficiency. Speech Text & Text Speech in All Languages application in which applying auto speech recognition technology is unable to catch up with the fast reader and low volume reader. The speed of the reading will affect the clearness of words that are spelled by the reader, and it gets harder for the application to understand the word so that it can convert into text. Based on the first cycle of the research, some students recorded their reading in hurry, so that the application converted the wrong words to the text that showed up in the application sheet. The condition improved when the researcher made sure that the students' reading speed was at normal speed with correct punctuation. The reader must stop at the end of every sentence and give a chance to the application to convert the sentences before continuing to next sentences. The application works very well when the reader controls their reading speed.

CONCLUSION

This research support theories that is stating that technology will help learners to get significant improvement and finding a solution for both English teachers and students for better teaching and learning experience. The use of Auto speech recognition technology that successfully adjust to audio feedback activity through Speech Text & Text Speech in All Languages application has proven as an effective strategy as shown by the ten grader students of SMAN 1 Lawang language major who are able to consistently increase the classroom average score of pronunciation skills.

More than a half of the class is also showing consistent improvement as well as the improvement of the class average score that recorded as eighty two in the first attempt, increased to eighty four in the second attempt of recording, then a big increase in the third attempt as much as eighty eight, and it reached ninety in the fourth attempt. The use of the application also minimize time to finish one topic or material and also provide an objective scoring method of students' speaking skills.

This research is also in line with previous researches that have been done in 2017 both for Chinese students and Iraqis students. Both of the researches finding were auto speech recognition program enable to improve the students pronunciation, increase the students' score, and provide an opportunities to self-learning for shyer students so that they will not stressed out with unnecessary jittery feeling, while this research is identical.

Therefore, the use of auto speech recognition is proven to be a helpful media to support the English classroom activity, especially in SMAN 1 Lawang.

REFERENCES

Hrehova, D. (2010). The Significance of Communication and Language Competences for Employment and Career Development. *Human Resources Management & Ergonomics Volume IV*, 2.

Kemmis, S., & McTaggart, R. (1998). *The Action Research Planner. (Third Edition).* Waurn Ponds: Deakin University Press.

M. Li, et al., (2017). Improving English Pronunciation via Automatic Speech Recognition Technology.International Symposium on Educational Technology.

McKay, S. (2002). *Teaching Engish as an International Language.* Oxford: Oxford University Press.

Sidgi, L. F. (2017). The Effect ofAutomatic Speech Recognition EyeSpeak Software on Iraqi Students' English Pronunciation:A Pilot Study. *Advances in Language and Literary Studies*, 48-54.