Advances in Language and Literary Studies
ISSN: 2203-4714
Vol. 8 No. 1; February 2017
Australian International Academic Centre, Australia
Flourishing Creativity & Literacy

The Usefulness of Automatic Speech Recognition (ASR) Eyespeak Software in Improving Iraqi EFL Students' Pronunciation

Lina Fathi Sidig Sidgi (Corresponding author)
Universiti Utara Malaysia, College of Arts and Sciences, 06010 Sintok, Kedah, Malaysia
E-mail: lheart41@yahoo.com

Ahmad Jelani Shaari
Language Department, School of Education and Modern Languages, Universiti Utara Malaysia, College of Arts and Sciences, 06010 Sintok, Kedah, Malaysia
E-mail: jelani@uum.edu.my

Doi:10.7575/aiac.alls.v.8n.1p.221
URL: http://dx.doi.org/10.7575/aiac.alls.v.8n.1p.221
Received: 14/12/2016
Accepted: 13/02/2017

Abstract

The present study focuses on determining whether automatic speech recognition (ASR) technology is reliable for improving English pronunciation to Iraqi EFL students. Non-native learners of English are generally concerned about improving their pronunciation skills, and Iraqi students face difficulties in pronouncing English sounds that are not found in their native language (Arabic). This study is concerned with ASR and its effectiveness in overcoming this difficulty. The data were obtained from twenty participants randomly selected from first-year college students at Al-Turath University College from the Department of English in Baghdad-Iraq. The students had participated in a two month pronunciation instruction course using ASR Eyespeak software. At the end of the pronunciation instruction course using ASR Eyespeak software, the students completed a questionnaire to get their opinions about the usefulness of the ASR Eyespeak in improving their pronunciation. The findings of the study revealed that the students found ASR Eyespeak software very useful in improving their pronunciation and helping them realise their pronunciation mistakes. They also reported that learning pronunciation with ASR Eyespeak enjoyable.

Keywords: automatic speech recognition, ASR, pronunciation instruction, English as a foreign language, EFL

1. Introduction

Technological evolution must develop at a fast pace to operate in today's progressive work environments. Certain concepts in the current technological progression still require some genuine self-reliant systems. Such systems can be created successfully by inculcating the way artificial intelligence can communicate in a real way that speaks like humans (Beelders & Blignaut, 2011). Automated speech recognition (ASR) technology has seen similar technical developments and authentic progress in inventions that are beneficial for specialised sets of users such as language learners to help them improve their spoking skills in the target language. ASR is a leading technology that allows humans to interact with data-processing programs through vocalisation. It resembles general conversation among humans; thus, ASR is currently one of the most intelligent developments in technology.

Learning English pronunciation is considered a difficult task especially for foreign language learners. In Iraq, learners of English find it difficult to masterin pronunciation (Hassan, 2014(. The factors that influence the accuracy of English pronunciation include mother tongue interference (Carter & Nunan, 2001) where the sound system of the tareget language and the mother tongue language are very different or nonexistent. There are certain English sounds, such as /p/, /v/, /ʧ/, /ŋ/, /ʒ/ (Odisho, 2005) and /e/, /ɔ:/, /ɒ/, /ɜ:/, /ə/, and /ɑ:/ that are not found in the Arabic sounds system. Arab learners face problems pronouncing sounds that they are not familiar with (Homeidan, 1984). Therefore, using the technology of ASR may help them improve their pronunciation. Using ASR software gives the students the opportunity to hear and practice the sounds of the English language. Also, it provides the students with individual practise, feedback, and immediate correction. The software used in this study to measure the usefulness of ASR in improving Iraqi students' English pronunciation was Eyespeak. This ASR software has some features that may help students improve their pronunciation that include: (a) listening to the native speaker pronunciation, (b) allowing the students to record their own pronunciation and compare it with the native speaker pronunciation, (c) providing immediate feedback and showing the student's mistakes with an explanation of the mispronunciation, (d) delivering a variety of feedback forms such as showing an animated sound wavelength that compares the student's pronunciation with the native speaker pronunciation and an animated side figure of human speech organs, which shows the place of articulation for each sound.

2. Basic Concept of Automatic Speech Recognition

In the fields of electrical engineering and computer science, speech recognition, or speech-to-text, has great importance in translating expressed words or content into textual matter. Automatic speech recognition is the process whereby human speech is interpreted by a computer (Forsberg, 2003) through the process of transferring the signals that are generated by the human vocal system into words (Jurafsky & Martin, 2000). The acquired speech is first digitised, confirmed against a dictionary, and then converted and, if required, displayed as typed text (Freedman, 1998). ASR allows electronic devices to determine phrases spoken by an individual learner via a microphone or a telephonic device in order to convert it into textual content. Its usage on a daily basis enhances its quality and allows for its use in distinct applications (Cucchiarini, 2009). According to Stark, Whittaker, and Hirschberg (2000), ASR has shown 100% quality and accuracy in apprehending all languages, but only when inputting fluently verbalised speech. ASR is unable to comprehend speech when an individual is unable to make clear and fluent statements. Thus, with the intent of understanding broader forms of speech, researchers are attempting to improve ASR by configuring its ability to recognise a much larger set of vocabularies. For example, ASR could be applied to learning foreign languages if it was able to interpret less than exact pronunciations.

ASR is used for Eyespeak software, which interprets English pronunciation by foreign students. It is especially useful for students who are trying to learn English as a second language and who tend to show a prior tendency of improvising pronunciation (Kim, 2006). Natural language processing is the most progressive method for interpreting language of all the currently developed ASR technologies. It allows individuals to have real conversations with intelligent machines, but there is still some room for improvement to reach 96–99% accuracy. Further, the highly advanced systems of the Siri interface for the iPhone aids individuals in having open-ended conversations that imitate real conversation, as Siri is not restricted to a limited set of words (Demenko, Wagner, & Cylwik, 2010). Directed dialogue conversation is another form of ASR that allows for a limited set of choices by an individual to converse with the machine interface. It offers narrowly outlined requests to individuals in order to acquire considerable knowledge that is beneficial for daily usage, for example, in many machine-driven, telephone-banking services along with other customer service interfaces.

3. Improving Students' English Pronunciation with ASR

ASR is a type of computer-assisted language learning (CALL) that allows non-native English learners to improve their pronunciation skills. The increasing use of speech technology has mostly been seen in the field of foreign language pedagogy (Strik & Cucchiarini, 2009). This has led to the evolution of CALL, which has resulted in some potent benefits. CALL has led teachers' constricted time periods by reducing the dependency of students on teachers. In addition, it has allowed users to work at their individual paces without stress and to observe their progressive acquisition of the subject matter. Finally, it has allowed for the continual access to other additive learning materials, such as visualisation and recordings, that play an important role in learning English as a foreign language.

CALL largely supports a creative outlook in the field of language teaching that focuses on informative constructivism in a multimedia-based environment collaborated on by students and instructors. In the past, pronunciation in English was largely ignored in language learning in favour of vocabulary and grammar, even though tutors believed that this had a harmful effect on pronunciation skills (Strik, 2009). However, later investigations clarified that language pronunciation has great importance in communication and improves the cognitive level of learners. In addition, training largely contributes to improving language pronunciation. Thus, computer-assisted pronunciation training (CAPT) was created to aid scholars in improving their English pronunciation skills using the technical measures of CALL.

4. Dimensions of ASR-Based CALL

CALL, which is composed of ASR traits, provides an optimum solution to pronunciation learning consisting of three distinct attributes: pedagogical requirements, audio-visual training sessions and speech technology.

4.1 Pedagogical Requirements

The concept of learning pronunciation is based on distinct methods of teaching (Wester, 2013). The overall measures of CAPT can be further categorised into three distinct components: input, output, and feedback.

4.1.1 Input

Input is the most important factor in successful language learning, and many studies have proven the benefits of inputs in pronunciation learning (Zavaliagkos, 2011). Therefore, students should be able to access many inputs in order to acquire many of the existing accessible target models of learning. However, learning should revolve around the needs and demands of the students and should facilitate the learner in the long term. Inputs should be presented in the format (i.e., written, oral, audio, visual) that corresponds with the individual's learning style.

4.1.2 Output

Output is a significant factor of CAPT that concentrates on the process of listening. Non-native English learners focusing on pronunciation should actively practise the English accent (Elimat & AbuSeileek, 2014). Outputs encourage speech production by creating a stress-free environment in which students are not hesitant to engage with the material. Outputs are used in communicative tasks mostly concerned with generating speech in the target language. When the students generate speech in the foreign language and practice to improve their pronunciation, they will maintain self-confidence in speaking the target language.

4.1.3 Feedback

Feedback is a debatable constituent of CAPT, and it has not been investigated much; thus, its effectiveness has not been specified. However, it is of special concern for students who wish to learn English as a second language, though its role has been deliberated (McCrocklin, 2014). Most commonly, feedback is given by teachers and is fundamentally known as repetition with change. It should not rely on the perceptions of scholars and should encourage learners to improve.

4.2 Audio-Visual Training Sessions

Another method of training via CAPT is through different measures of audio-visual devices (Neri, 2007). Audio-visual technicians are accountable for making a considerable number of setups, such as organised sets of conferences and seminars, to train individuals.

4.3 Teachers responsibility

It is the responsibility of educators to note any student's difficulty in delivering fluent speech in English (Zajechowski, 2014). Instructors must give feedback to improve learners' English pronunciation.

5. Effectiveness of ASR

A number of aspects including transfer of speech to text, scoring, error detection, error diagnosis, and feedback focus on ensuring that ASR is effective for educational purposes as a way to achieve objectives. ASR allows for the appropriate accessibility of those who are deaf or hard of hearing to improve their pronunciation, as pronunciation development is dependent on both oral and listening skills (Demenko, 2009). Traditional methods of pronunciation development are complex and enhance the overall costs of programs; ASR allows for cost reduction since it is automated.

The searchable text capability is a key benefit for pronunciation development through ASR. Diverse standards of pronunciation are well maintained through ASR as it allows for authentic interaction with the target language (Strik & Cucchiarini, 2009). ASR also promotes collective group work, which helps to encourage interaction among students. Pronunciation improvement is greatly dependent upon interactions with others because it is associated with oral skills. ASR allows students to participate in open-ended learning through diverse activities and to learn about the target language and its social contexts of use. The exploration of personal and societal goals can improve through the use of ASR (Zavaliagkos, 2011).

The success of ASR is dependent on the computers that provide support to the pronunciation modules. The computer is a tool to be used by students in order to properly utilise certain information. It also allows for language practice through online and software programs. Computers have a great capability to improve language learning and are advantageous to improving pronunciation (Wester, 2013). The appropriate analyses of sound patterns and foreign accent models are also significant because they allow the system to provide an outcome in the form of text. The use of ASR in the classroom is convenient and can save time for the teacher by providing the opportunity for all the students to practise pronunciation.

ASR can also be used in automated telephone lines. Therefore, pronunciation can also be advanced using Siri, an ASR program that is also dependent on pronunciation (Zavaliagkos, 2011). Siri only accepts a command if the pronunciation of words are accurate allowing students to use Siri to improve their pronunciation skills. When used for pronunciation training, ASR is a tool that allows students to practice at their own speed by receiving feedback from the words recognised by ASR.

Although ASR has been criticised for its low rates of accurate recognition for non-native speakers of the target language, this has been steadily improving. In addition, text assistance could be provided for sustainable development of ASR function. Researchers have been working continually to improve ASR's accuracy in evaluating spoken language; thus, ASR programs geared towards non-native speakers have reasonably high levels of accuracy (Strik & Cucchiarini, 2009).

6. Using ASR to Teach Pronunciation

Learning correct pronunciation is a critical aspect for university-level English learners. A key issue faced by students is the interpretation of foreign accents. To remedy this, ASR is being used effectively by teachers to spread awareness about the importance of correct pronunciation. ASR opens up new possibilities for training conversational skills and for adding new features such as listening to self-pronunciation production, listening to English native-speaker models, looking at animated sound production graphics, diagnosing student errors, and providing feedback; all of these features will support the teaching environment. However, it also requires the support of a CALL system (Demenko, 2009). With improved focus on pronunciation, self-directed learning can be promoted effectively within the classroom.

Education systems that incorporate ASR modules have better student interaction and identification of the issues learners face. With the assistance of ASR modules, immediate feedback can be provided to learners so that errors in pronunciation can be remedied effectively. Currently, a number of commercial institutions use ASR technology to teach second language pronunciation (Cucchiarini, 2009). In order to effectively use ASR technology, the use of CALL applications is also significant. ASR-based systems allow students to actively participate in oral skills modules. Through feedback, mistakes can also be identified. In this respect, the development of ASR modules is significant because it allows for the recognition of errors and score of non-native speech.

In addition, ASR allows for effective pronunciation training in the classroom with a teacher. It offers individual practice on the computer so that better support can be provided for pronunciation practise. It is necessary to ensure that lecture notes and overall student interaction levels are stored in log files during individual sessions (Strik, 2009). Teachers also

need to provide supervision to ensure that the learning standards are being well maintained and student problems are remedied. It has been noted that teacher presence is beneficial because it helps in reducing the anxiety of technophobic learners (Kommissarchik & Komissarchik, 2000). Overall, teaching programs have been found to be beneficial as they provide a certain degree of extrinsic motivation (Murray, 1999).

Moreover, effectively designed exercises provide support to students and help them improve their oral skills. Therefore, designed exercises will be beneficial for better pronunciation learning so that educator goals and objectives can be met. Pronunciation errors can be overcome using ASR (Kim, 2006), but frequent, persistent and reliably autocratic techniques must be used. Focus on pronunciation in education systems should be advanced in order to accomplish certain goals and objectives by having successful language learners who can communicate comprehensibly in the target language. Pronunciation can be improved by focusing on both segmental and suprasegmental aspects; for example, speech quality can be advanced by focusing on sounds, word stress, and sentence stress (Cucchiarini, 2009). Pronunciation training through ASR allows for the real-time development and feedback.

7. Data Collection Technique

This experiment study used three data collection techniques: testing, questionnaire, and interview. However, for this article, only the data collected from the questionnaire will be discussed.

The questionnaire responses were collected after the experiment procedure of two months of pronunciation instruction with the use of ASR (Eyespeak) software. The students had no prior experience in using this or any other ASR software. Twenty subjects were selected randomly from the first-year college students at AL-Turath University College from the Department of English in Baghdad, Iraq. The teaching process included the same teaching material of the traditional class with the addition of ASR (Eyespeak) software. The students had to attend three 45 minute classes per week, which included one class using traditional teaching methods and two classes using ASR (Eyespeak) software. At the end of teaching procedure, all the students were given a questionnaire. The questionnaire intended to measure the usefulness ASR (Eyespeak) software in improving the students' pronunciation. The students gave their opinions on a 5-point Likert scale (strongly agree, agree, neutral, disagree, and strongly disagree) questionnaire that consisted of 19 close-ended questions.

8. The Findings

Descriptive analysis was used to analyse the students' responses to the questionnaire. The findings of the study revealed that most of the students responded positively towards the usefulness of ASR (Eyespeak) software in improving their English pronunciation; 65% of the students reported that using the software improved their pronunciation and helped them recognise mispronounced English words. Fifty-five percent of the students said that practising similar sounds and comparing their own recorded pronunciation with the native English pronunciation was very useful in improving their pronunciation. Fifty percent of the students reported that practising pronunciation several times with the use of ASR (Eyespeak) raised their awareness of the correct pronunciation.

The students' responses towards the intention of using Eyespeak in pronunciation class showed that 55% of them liked using it, and 50% of them liked practising pronunciation with the software.

As part of how the students perceived the ease of Eyespeak usage in their learning pronunciation process, the students' responses showed that 60% found using Eyespeak software enjoyable; 60% said that using the software did not require much effort; 55% found that practising with the software made them feel more comfortable and not afraid of making pronunciation mistakes; 45% of the students reported that feedback graphics helped them realise where they mispronounced; and 40% of them said that listening to native English pronunciation helped them in pronouncing difficult words.

Very few negative responses were recorded: 5% felt that the visual graphic feedback was not helpful in showing mispronunciations; 5% said that they experienced difficulties in using the software; and 5% felt that the software required too much mental effort.

The overall evaluation of the ASR (Eyespeak) software indicates that it is very useful in learning pronunciation. Most of the students found that using ASR (Eyespeak) software improved their pronunciation and made them realise their mistakes. Also, positive responses towards using the ASR (Eyespeak) software indicate that students felt more comfortable and that they found the lessons enjoyable. In a similar study, Nylén (2009) found that the students' attitudes were positive towards using Eyespeak as a tutor and that their rating scores indicated a significant improvement in the students speaking skills. On the contrary, few negative responses were made by the students towards the software. The negative responses indicated that the visual graphic feedback was not helpful in showing pronunciation mistakes.

Similarly, a study by Witt (2012) mentioned that Eyespeak exercise feedback gives the student’s score on the entire pronounced word and not specifically on the phoneme-level error, which may not help them understand their pronunciation errors. Witt (2012) stated that Eyespeak software has the useful feedback features of displaying the tongue position of student's sound production compared with the tongue position of native speaker's sounds production. Similarly, a study by Ai (2013) indicates that most ASR software provides the sound wavelength form of feedback as compared with native speaker pronunciation. But it does not show how to improve their curves (sound wavelength) to match the native speaker pronunciation curve. Therefore, it leads the students to try several times to produce the accurate pronunciation, which may lead to fossilisation (Eskenazi, 1999). In order to solve that problem, Ai (2013) suggested that using forced alignment might help students recognise the wrong phoneme to know when to increase or

reduce energy, loudness, or pitch as it used in Eyespeak.

9. Conclusion

In conclusion, ASR plays a very crucial role in enhancing the learning of individuals who are trying to improve their English pronunciation skills. In this paper, pronunciation was referred to as a subsidiary deliberation of foreign language learning and was bifurcated into five pivotal divisions. This paper defined the significance of ASR in contributing to improvement in the pronunciation of English learners and specified the different dimensions of ASR to measure its effectiveness. Finally, the several roles of ASR in teaching pronunciation were covered. This paper suggests that the ASR method is found to offer a great opportunity in teaching and more useful learning pronunciation than traditional instruction. The educational environments with the use of ASR in the classroom are highly motivating and reduce anxiety in learning English pronunciation.

References

Ai, R. (2013). Recent advances in natural language processing: Perceptual feedback in computer assisted pronunciation training: A survey. [Online] Available: http://aclweb.org/anthology/R/R13/R13-2001.pdf (February 18, 2017).

Beelders, T.R., & Blignaut, P.J. (2011). The usability of speech and eye gaze as a multimodal interface for a word processor. INTECH Open Access Publisher. https://doi.org/10.5772/16604

Carter, R., & Nunan, D. (2001). *The Cambridge guide to teaching English to speakers of other languages.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667206

Cucchiarini, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication*. 51(10), 845–852. https://doi.org/10.1016/j.specom.2009.05.007

Cylwik, N., Wagner, A., & Demenko, G. (2009). *The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System*, Proceedings of SLaTE Workshop on Speech and Language Technology in Education, Wroxall Abbey Estate, Warwickshire, 85–88.

Demenko, G., Wagner, A., & Cylwik, N. (2010). The use of speech technology in foreign language pronunciation training. *Archives of Acoustics*, *35*(3), 309–329. https://doi.org/10.2478/v10168-010-0027-z

Elimat, A. K., & AbuSeileek, A. F. (2014). *Automatic speech recognition technology as an effective means for teaching pronunciation.* [PDF]. Available: http://journal.jaltcall.org/articles/10_1_Elimat.pdf (December 5, 2016).

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

Forsberg, M. (2003). Why Is Speech Recognition Difficult? Chalmers University of Technology. [Online] Available: http://www.speech.kth.se/~rolf/gslt_papers/MarkusForsberg.pdf (February 20, 2017)

Freeman, D. (1998). 'Catch[ing] the nearest way': Macbeth and cognitive metaphor. In Culpeper, J., Short, M. & Verdonk (eds.), Exploring the Language of Drama: From text to context. London: Routledge, 96-111.

Hassan, E. (2014). Pronunciation problems: A case study of English language students at Sudan University of Science and Technology. *English Language and Literature Studies*, *4*(4). http:/doi:10.5539/ells.v4n4p31

Homeidan, A. H. (1984). *Utilizing the Theory of articulatory settings in the teaching of English pronunciation to Saudi students learning English as a second language.* (Doctoral dissertation), King Fahd Public Library, Jeddah, Saudi Arabia.

Jurafsky, D. & Martin, J. (2000). *Speech and language processing* (1st ed.). Upper Saddle River, N.J.: Prentice Hall.

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

Kommissarchik, J., & Komissarchik, E. (2000). Better accent tutor analysis and visualization of speech prosody. *Proceedings of InSTILL*, 86–89.

McCrocklin, S. M. (2014). The potential of automatic speech Recognition for fostering pronunciation learners' autonomy. [Online] Available: http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=4909&context=etd (December 5, 2016).

Murray, G. L. (1999). Autonomy in language learning in a simulated environment. *System* (27), 295–308. http://doi.org/10.1016/S0346-251X(99)00026-3

Neri, A. (2007). The pedagogical effectiveness of ASR-based computer assisted pronunciation training. [Online] Available: http://hstrik.ruhosting.nl/wordpress/wp-content/uploads/2013/02/Neri-PhD-thesis.pdf (December 5, 2016).

Nylén, P. (2009). Learning English with the use of ICT: An action research study on students' attitudes. Växjö University. DIVA Academic Archive. [Online] Available: http://www.diva-portal.org/smash/get/diva2:246166/FULLTEXT01.pdf (February 18, 2017).

Odisho, E. (2005). Techniques of teaching comparative pronunciation in Arabic and English. New Jersey: Gorgias Press LLC.

Stark, L., Whittaker, S., and Hirschberg, J. (2000). ASR satisficing: the effects of ASR accuracy on speech retrieval. In *Proceedings of International Conference on Spoken Language Processing*.

Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication,* 51(10), 853–863. https://doi.org/10.1016/j.specom.2009.03.003

Strik, H., & Cucchiarini, C. (2009). Modeling pronunciation variation for ASR: A survey of the literature. *Speech Communication.* 29(2), 225–246.

Wester, M. (2013). Pronunciation modeling for ASR–knowledge-based and data-derived methods. *Computer Speech & Language.* 17(1), 69–85.

Witt, S.M. (2012). Automatic error detection in pronunciation training: Where we are and where we need to go. *In International Symposium on Automatic Detection of Errors in Pronunciation Training, Stockholm, Sweden*.

Zajechowski, M. (2014). Automatic speech recognition (ASR) software – An introduction. [Online] Available: http://usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/ (December 5, 2016).

Zavaliagkos, G. (2011). Stochastic pronunciation modelling from hand-labelled phonetic corpora. *Speech Communication,* 29(2), 209–224. https://doi.org/10.1016/S0885-2308(02)00030-X