Research article

The Potential of ASR for Improving English Pronunciation: A Review

Rahmati Putri Yaniafari*, Viga Olivia, and Suharyadi

Universitas Negeri Malang

ORCID
Rahmati Putri Yaniafari: https://orcid.org/ 0000-0001-8575-6947

Abstract.
To pronounce well is a complex task, requiring students not only to possess knowledge of the appropriate sounds in a given context, but also to learn to use their vocal apparatus to make those sounds, equipped with extensive practice and feedback. Students in these situations require autonomous monitoring experiences to receive tailored feedback. One of the technological tools learners can use to improve their pronunciation is Automatic Speech Recognition (ASR). This provides learners with individual practice and feedback to assist them to accomplish their language goals. This study examines the database of research on the use of ASR in pronunciation instruction and learning available on Google Scholar, Springer Link, Education Resources Information Center (ERIC), Taylor & Francis Online and Directory of Open Access Journal (DOAJ). To help the process of identification, some procedures and criteria were employed. The results revealed that ten articles met the eligibility criteria. The procedures of utilizing ASR to improve students' pronunciation competency are then discussed in this study.

Keywords: teaching speaking, teaching pronunciation, Automatic Speech Recognition

Corresponding Author: Rahmati Putri Yaniafari; email: yaniafari.fs@um.ac.id

Published 28 March 2022

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the iNETAL Conference Committee.

1. Introduction

One of the main requirements for students' language proficiency is to have securing pronunciation [1]. This means that students should have an ability to know the sound of a word which is commonly different from its written. Pronunciation becomes a a challenging task as it requires students not only to have knowledge of the relevant sounds in a given context, but also to learn how to pronounce those sounds using a vocal device, under extensive practice and feedback [2]. Pronunciation is defined as the act of creating a speech sound that includes intonation articulations, vowels, accent inflection and formation, in relation to the accuracy or acceptance of utterance [3]. This complexity is often overlooked as it allows them to deal with phonetic and phonological competence. Instead, learners emphasize fluency and communication skills over phonetic accuracy as they have experienced less training on them.

Under these circumstances, students need independent monitoring experience to get individualized feedback. This is an alternative way so that they are no longer a reliant on the teacher who considered as the main facilitator. Language learning appears to be most efficient when the teacher constantly monitors progress to guide remediation or advancement [4] . However, in teacher-fronted lessons there is generally not enough time for sufficient practice and feedback on speaking performance, while traditional language lab tools usually do not provide the feedback required [5]. Shortly, due to time constraints in class, providing such training and correcting all student errors is very difficult to achieve. This encourages the learning process to take advantage of the technology that has been used this year to assist learning. The rapid development of technology has opened up an attractive and convenient new arena for language learning. Particularly, the mobile phone, along with its ubiquitousness and mobility, offers potential advantages in foreign language learning [6].

With the advancement of information technology, variety of speech tools have emerged to help learners work on their pronunciation. Automatic Speech Recognition (ASR) is one of which that help learner reach the language target by offering individual practice and feedback. ASR is a leading technology that allows humans to interact with data-processing programs through vocalization [7]. The most advanced systems incorporating automatic speech recognition (ASR) technology can provide feedback at the sentence, word, or text level [8]. ASR provides authentic material, such as native speakers' pronunciation of the target language, and at the same time allows the students to listen to and practice their pronunciation in an enjoyable setting; it also gives each individual learner immediate corrections and feedback, which is difficult to achieve in class with a large number of students [9]. A stress-free environment promised by this method encourages students to participate more as autonomous learners. It is very essential as it could give significant impact on the learners' pronunciation skill.

Therefore, this study is aimed to review the recent studies of the use of Automatic Speech Recognition (ASR) in teaching/ learning pronunciation.

2. Recent studies on the Use of ASR to Improve Pronunciation Competence

In the process of identification of relevant studies, a search for articles, conference papers and thesis was done thoroughly through some research databases, which are Google Scholar, Springer Link, Education Resources Information Center (ERIC), Taylor & Francis Online and Directory of Open Access Journal (DOAJ). Several search terms

used include “Automatic Speech Recognition”, “Pronunciation”, and “Automatic Speech Evaluation”, combined with “speaking” and “EFL”. To get the most recent studies, the articles searched were filtered to those published from 2016 to 2020. The following criteria of eligibility were also set in view of the research questions.

1. The paper aims at proposing, implementing, or reviewing the use of Automatic Speech Recognition in teaching/ learning pronunciation
2. The paper is written in English
3. The paper is published between 2016 and 2020
4. The paper is conducted in the context of EFL/ESL

The article identification process (See Figure 1) revealed that there are 10 articles meeting the criteria of eligibility. The initial search through the aforementioned research databases shows a great number of papers (n=85). However, after the filtering process, most of the articles were excluded due to not meeting the requirement of eligibility.

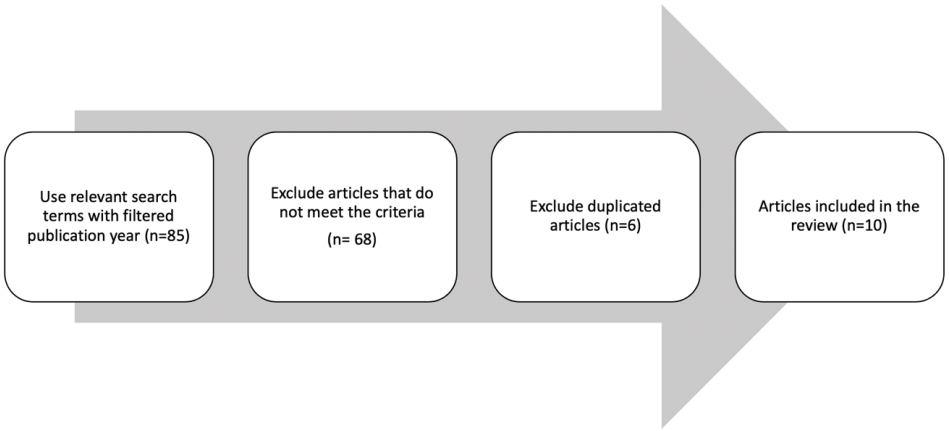

Figure 1: The Process of Articles Identification.

The studies on the use of ASR in EFL setting were conducted on different levels of education. Out of 11 studies, 2 were investigating the ASR use in EFL classes for a company employees, 6 were in tertiary level, and 2 others in high school. Table 1 shows the detail of the summary of the studies.

3. ASR Applications Used

From the 10 studies reviewed, some recommend method in using ASR in teaching pronunciation. Some of the ten studies that looked into the use of ASR in ESL teaching and learning employed web-based ASR, while others used software or smartphone

Table 1: Summary of the Studies.

No.	Author	Year	Research Subject	ASR Used/ Reviewed	The Focus of the Investigation
1.	Lara Wallace	2015	University Student	Google Web Speech	Proposing an alternative strategy to identify personal pronunciation problem using ASR
2	Saandia Ali	2016	Graduate and Undergraduate students	Various Smartphone Commercial Apps and CAPT Software	Reviewing existing ASR tools as a preliminary step toward ASR software development for CAPT
3	Shannon M. McCrocklin	2016	University students joining an Advanced ESL Listening course	Windows Speech Recognition and voice search on smart phones	Investigating the effect of using ASR on students' autonomous learning beliefs and practices
4	Lina Fathi Sidig Sidgi	2017	First year university students of English Department	EyeSpeak Software	Investigating the effectiveness of the ASR software in improving students' pronunciation
5	Xiaobin Liu, et al.	2018	Miidle School Students	Fluent English learning app	Trying to see whether the application facilitates the students' pronunciation learning
6	Xiaobin Liu	2019	Freshmen University Students	iFlytek Voice Input (IVI) application	Investigating whether the application helps in improving students' English pronunciation
7	Katerina Evers1 and Sufen Chen	2020	Employees of International Sports Company joining Business English Course	"Speechnotes – Speech to Text"	Investigating the impact of various teaching methods on the pronunciation performance of students with various learning styles are being investigated.
8	Katerina Evers & Sufen Chen	2020	Employees of International Sports Company joining Business English Course	"Speechnotes – Speech to Text"	Investigating whether teamwork and feedback affect learners' views towards ASR
9	Caiyun Liu	2020	College students	Not mentioned	Exploring the phoneme association and phoneme recognition results in English
10	Yuting Yuan and Xiaobin Liu	2021	3^{rd} graders of Senior Highschool	"Oral English Drill & Test" App	investigating the effects of ASR in assisting reading aloud practices on students' pronunciation

apps, and the remainder did not mention the tool explicitly. The web-based ASR applications used include Google Web Speech [10] and Speechnotes [11], [12]. When this article is written, a search on google with "Google Web Speech" key words led to a

page called ”Speech to Text by Google,” which offered us a free trial of the service as well as a price for the premium version. The page also provides a demo that students can access (See Figure 2). The second web-based ASR application, Speechnotes, also has both the basic account and the premium offer. When we go to *https://speechnotes.co*, however, we can immediately see a clear explanation on how the ASR works, including how to dictate, punctuate, and proofread (See Figure 3).

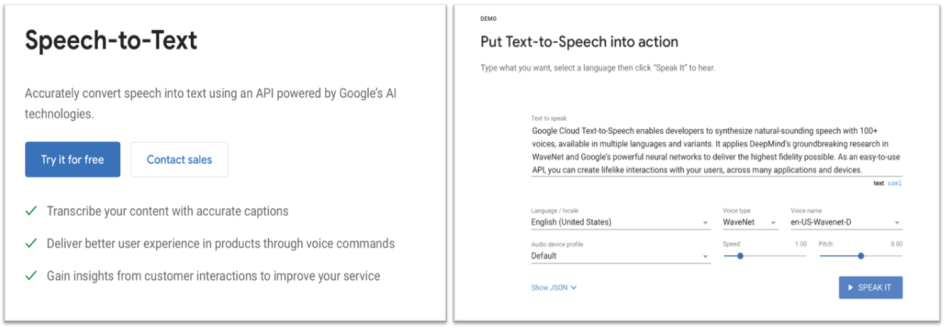

Figure 2: The Google Cloud Text-to-Speech Webpage.

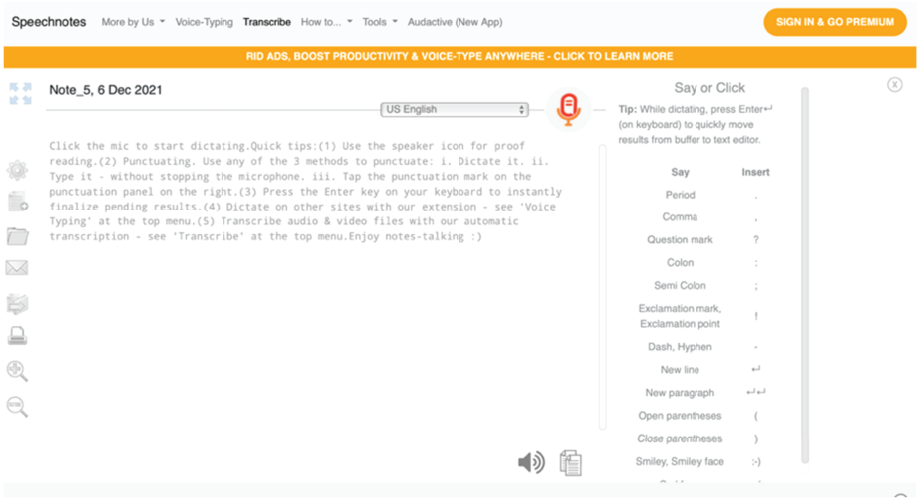

Figure 3: The Google Cloud Text-to-Speech Webpage.

The rest of the studies utilized ASR software or smartphone applications, which are Windows Speech Recognition [13], EyeSpeak Software [9], iFlytek Voice Input [14], Oral English Drill & Test [15], and Fluent English learning app [16]. Among those software and applications, only Windows Speech Recognition used by McCrocklin which is accessible to the author when this paper is written. It can be found in windows computer without and does not require further installation. The user merely need to verify the settings on their computer’s ”control panel,” which they may do without having to go online. This

feature allows voice commands and text dictation in electronic documents (See Figure 4).

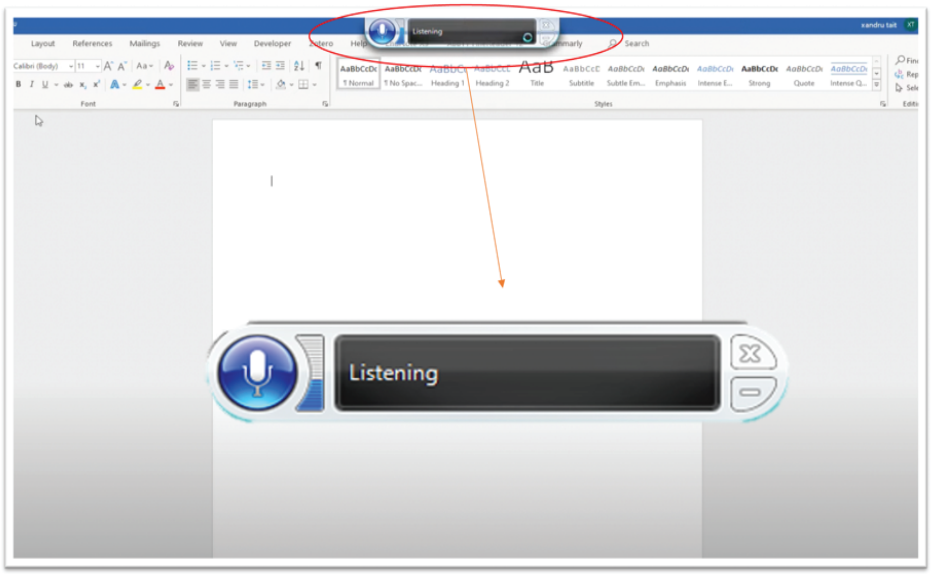

Figure 4: Windows Speech Recognition.

All the ASRs used speech to text procedure, which transcribes any utterances we make using the connected microphone. EyeSpeak, on the other hand, based on a review by Tao [17], is designed to facilitate students in learning to produce English phonemes accurately. It has animated images, sound recordings, graphic articulatory displays, and record/playback capabilities, according to Tao's software evaluation (See Figure 5).

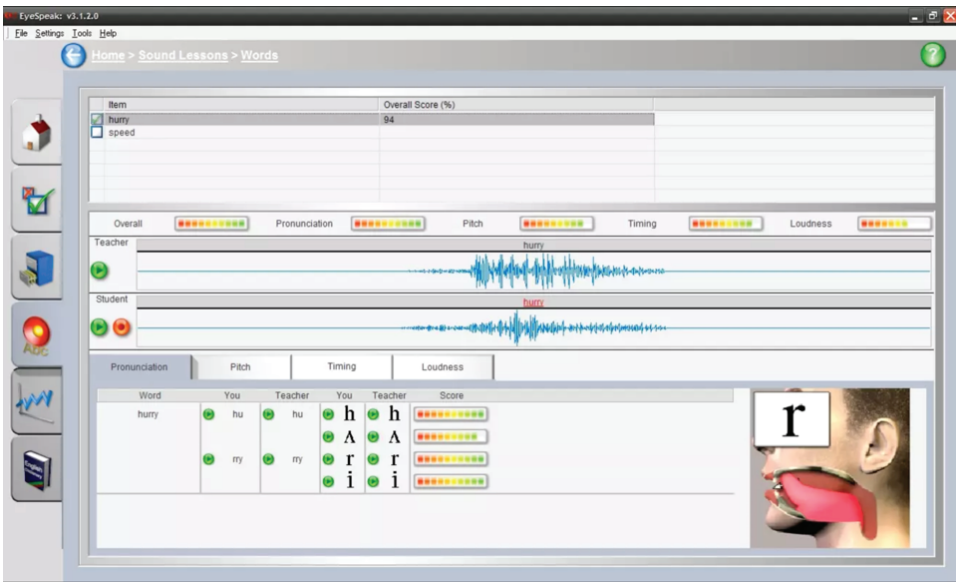

Figure 5: EyeSpeak English User Interface [18].

4. Proposed Procedure in Using ASR in Teaching Pronunciation

Wallace recommends a method for detecting pronunciation issues using ASR [10]. The process began with the students recording their speech and ASR transcribing it. The pupils then rectify the incorrect transcriptions by highlighting them. Next, they re-listen to their recording and analyze the highlighted words that appear to be mispronounced. Liu further categorized the pronunciation errors based on the recognition results into three types: (1) *Missing phoneme* (phonemes that do not exist in the transcription, but is silent), (2) Misreading phonemes (phonemes that do not exist in the transcription and not a missing phoneme), and (3) Added Phonemes (Extra phoneme in the transcription)[2].

However, Evers & Chen's study of the impact of various teaching methodologies on the pronunciation performance of students with various learning styles suggests that ASR feedback has limits in terms of assisting pronunciation improvement [11]. The findings suggest that combination of ASR feedback and peer-assistance as the most effective approach compare to the use of ASR alone or teacher's feedback. Moreover, the results are applicable for both students with visual and verbal learning style.

5. Conclusion

The analysis of the database shows that the studies on the use of ASR were mostly conducted in classrooms with adult learners (Out of 11 studies, 2 were investigating the ASR use in EFL classes for a company employee, 6 were in tertiary level, and 2 others in high school). The studies revealed that ASR is beneficial in assisting pronunciation learning. However, although practice using ASR alone is effective [7, 9, 10, 13, 15, 19], a combination of ASR pronunciation activities with peer instruction may yield even better results as individual work may frustrate language learners by highlighting the mistakes in their speech that they cannot fix on their own.

References

[1] Gilakjani AP. A study of factors affecting EFL learners' English pronunciation learning and the strategies for instruction. International Journal of Humanities and Social Science. 2012;2(3):119–128.

[2] McCrocklin SM. Pronunciation learner autonomy: The potential of Automatic Speech Recognition. System. 2016;57:25–42. https://doi.org/10.1016/j.system.2015.12.013

[3] R. M. Dauer, Accurate English: A Complete Course in Pronunciation. Hoboken: Regents/Prentice Hall, 1993.

[4] Ali S. Towards the development of a comprehensive pedagogical framework for pronunciation training based on adapted automatic speech recognition systems. Paper presented at: EURO CALL 2016: Conference on Computer Assisted Language Learning (CALL); August 24—27, 2016; Lemesos, Cyprus.

[5] van Doremalen J, Boves L, Colpaert J, Cucchiarini C, Strik H. Evaluating automatic speech recognition-based language learning systems: A case study. Computer Assisted Language Learning. 2016;29(4):833–851. https://doi.org/10.1080/09588221.2016.1167090

[6] Youn Ahn T, Lee S-M. User experience of a mobile speaking application with automatic speech recognition for EFL learning: Speaking app with ASR. British Journal of Educational Technology. 2016;47(4):778–786. https://doi.org/10.1111/bjet.12354

[7] Sidgi LF, Shaari AJ. The usefulness of automatic speech recognition (ASR) eyespeak software in improving Iraqi EFL students' pronunciation. Advances in Language and Literary Studies (ALLS). 2017;8(1):221–226 . https://doi.org/10.7575/aiac.alls.v.8n.1p.221

[8] Elimat AK, AbuSeileek AF. Automatic speech recognition technology as an effective means for teaching pronunciation. Japan Association for Language Teaching Computer Assisted Language (JALTCALL). 2014;10(1):21–47. https://doi.org/10.29140/jaltcall.v10n1.166

[9] Sidgi LFS, Shaari AJ. The effect of automatic speech recognition eye speak software on Iraqi students' English pronunciation: A pilot study. Advances in Language and Literary Studies (ALLS). 2017;8(2):48–54. https://doi.org/10.7575/aiac.alls.v.8n.2p.48

[10] Wallace L. Using Google web speech as a springboard for identifying personal pronunciation problems. Paper presented at: The 7th Annual Pronunciation in Second Language Learning and Teaching Conference; October 15—17, 2015; Dallas, USA.

[11] Evers K, Chen S. Effects of automatic speech recognition software on pronunciation for adults with different learning styles. Journal of Educational Computing Research. 2020;59(4):669–685. https://doi.org/10.1177/0735633120972011

[12] Evers K, Chen S. Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. Computer Assisted Language Learning. 2020;59(4):686–712. https://doi.org/10.1080/09588221.2020.1839504

[13] McCrocklin SM. Pronunciation learner autonomy: The potential of Automatic Speech Recognition. System. 2016;57:25–42. https://doi.org/10.1016/j.system.2015.12.013

[14] Liu X, Xu M, Li M, et al. Improving English pronunciation via automatic speech recognition technology. International Journal of Innovation and Learning. 2019;25(2): 126—140.

[15] Yuan Y, Liu X. An empirical study of the effect of ASR-supported English reading aloud practices on pronunciation accuracy. Paper presented at: The 5th International Conference on Technology in Education (ICTE 2020); August 19–22, 2020; Macau, China.

[16] Liu X, Zhu C, Jiao J, Xu M. Promoting English pronunciation via mobile devices-based automatic speech evaluation (ASE) technology. Paper presented at: The 11th International Conference on Blended Learning (ICBL 2018); 31 July—2 August 2018; Osaka, Japan.

[17] Tao R. EyeSpeak: Software review. Computer Assisted Language Instruction Consortium (CALICO) Journal. 2008;25(1):126–136.

[18] Alvaro T. *EyeSpeak English software review*.Jan. 2009. Accessed on: October 30, 2021. [Online]. Available: https://eyespeak-english.en.uptodown.com/windows.

[19] Liu C. Application of speech recognition technology in pronunciation correction of college oral English teaching. Paper presented at the 2020 International Conference on Multi-model Information Analytics (MMIA2020); March 5–6, 2020; Changzhou, China.