



# The Application of Automatic Speech Recognition Technology in English as Foreign Language Pronunciation Learning

Yujie Cai<sup>(✉)</sup>

Hainan Normal University, Hainan, China

A348098752484@163.com

**Abstract.** The development of ASR technology has made it a potential solution for EFL Pronunciation learning. This study seeks to answer the following research questions: Q1. Can ASR-based tools provide accurate and individualized automated feedback for mispronunciation? Q2. Is ASR training effective for EFL pronunciation learning? Q3. How to utilize ASR technology as part of the pronunciation learning process? The research results showed that properly adapted ASR technology can provide an individualized, stress-free, self-paced learning environment as well as opportunities for immediate, individualized feedback. A blended EFL pronunciation learning method is proposed and future research direction is pointed out in this study.

**Keywords:** ASR · APED · EFL · feedback · pronunciation · CAPT

## 1 Introduction

Pronunciation is essential in EFL because it is the foundation of speaking, and without a certain degree of pronunciation accuracy, others will be unable to understand the speakers' utterances. To pronounce well is a complex task, one of the technological tools learners can use to improve their pronunciation is Automatic Speech Recognition (ASR). It can provide feedback on mispronunciations and facilitate EFL learners to improve their speaking accuracy and proficiency through practice.

## 2 Theoretical Background

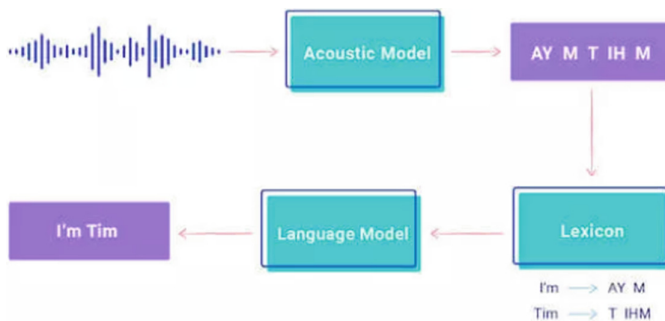
### 2.1 What is ASR and How ASR Works?

Automatic speech recognition (ASR), is a process of converting the speech into a sequence of words, using some algorithms which have been implemented as programs. [1] How ASR works? There are two main approaches to ASR, one is hybrid approach, another is end-to-end Deep Learning approach. The hybrid approach combines a lexicon model, an acoustic model and a language model to make transcription predictions. Despite plateaus in accuracy, the hybrid approach is widely used today (Fig. 1).

© The Author(s) 2023

I. A. Khan et al. (Eds.): HWESM 2023, ASSEHR 760, pp. 356–360, 2023.

[https://doi.org/10.2991/978-2-38476-068-8\\_44](https://doi.org/10.2991/978-2-38476-068-8_44)



**Fig. 1.** The hybrid approach model [2].



**Fig. 2.** End-to-end Deep Learning model [2].

A newer approach of ASR is end-to-end Deep Learning Approach. Without lexicon model and language model, it can directly map a sequence of input acoustic features into a sequence of words. The End-to-end Deep Learning models is more accurate than the traditional models (Fig. 2).

## 2.2 What is APED?

APED is an abbreviation of automatic pronunciation error detection. The APED system first gives a predefined utterance text, and the learner tries to pronounce this target text correctly. By accurately detecting the pronunciation errors and providing precise feedback on what is mispronounced, the APED system guides the learner to correct the pronunciation towards the target utterance and improve the speaking ability [3].

## 3 Literature Review

The research results of Adhan Kholis showed that ASR APP can improve the students' pronunciation skills and motivate the students to engage in learning to pronounce [4]. Thomas Dillon and Donald Wells' research results suggested that many Korean students were open to the incorporation of ASR in the EFL classroom and there was significant potential in incorporating ASR into EFL classroom [5]. Xiaobin Liu and Manfei Xu studied the application of ASR technology in English pronunciation correction. The research result showed that the ASR-based device can improve English pronunciation accuracy and provide a way for learners to improve their English pronunciation [6]. In-Seok Kim's research on the reliability of ASR software used to teach English pronunciation revealed that the present state of ASR technological development falls far below the desired level of accuracy [7]. Riikka Muhonen's research results showed that ASR can facilitate

assessment and reduce pupils' foreign language anxiety [8]. Ryan Spring found that the feedback from an ASR-based practice tool caused Japanese EFL learners to focus on their pronunciation and adjust it. The results proved that ASR-based pronunciation tools are meaningful in a wide variety of Japanese EFL contexts [9]. Reece Randall's research results showed that computer-assisted pronunciation training (CAPT) helped Korean EFL learners improve their pronunciation and motivation [10]. Wen-Hsin Chen investigated the effect of ASR on segmental development and explores learners' perception of ASR as a pedagogical tool. The results revealed that learners significantly improved the /æ/-/ɛ/distinction in production [11]. The research results of Muzakki Bashori proved that ASR assisted websites can improve Indonesian EFL learners' vocabulary knowledge and pronunciation skills [12]. Tomoko Takenouchi's research result showed that explicit pronunciation instruction and pronunciation practice using ASR were effective for Japanese adult EFL learners [13]. The research of William Gottardi revealed that the use of ASR technology is a way of giving learners the opportunity to focus on their specific pronunciation difficulties and receive personalized feedback while becoming more autonomous in their learning process [14]. Solène Inceoglu's study showed that the majority of the participants perceived ASR as useful for pronunciation practice [15]. The research results of Spring Ryan revealed that the ASR tool was found to be most useful for students who had a lower than 95% accuracy on their pretests and was most helpful for practicing consonant and vowel sounds [16]. Jonás Fouz-González explored the potential of EFP app to help EFL learners improve their pronunciation. Improvements in the learners' perception and production of the target features offered a clear measure of the learning potential of the app [17].

## 4 Conclusion

The answer of Q1 is: APED in ASR systems can help teachers with pronunciation assessment tasks, allowing for the simultaneous and fast assessment of several students, providing objective assessment and reducing teacher's working hours. The use of ASR tool in EFL pronunciation learning seems to be beneficial and practical. The answer of Q2 is ASR-based training had an overall positive impact on pronunciation learning. Because ASR systems provide low-anxiety learning environments and timely feedback, many commercial computer-assisted language learning (CALL) programs, have incorporated ASR technology to encourage autonomous language learning outside of class. The answer of Q3 is: based on the current situation of English pronunciation teaching in China, this paper suggests adopting a blended English pronunciation learning method of combining English pronunciation instruction in class and automatic speech recognition training after class. Although there are many ASR tools, there is little quality control to assess their rigor and effectiveness in pronunciation learning as self-learning resources. To solve this problem, we call for greater collaboration between pedagogic and technical experts when designing ASR training products. It should have a pronunciation syllabus with clear learning objectives or outcomes. It should also be able to make individualized pronunciation training plans according to students' errors. When ASR recognizes a pronunciation error, it should provide an audiovisual feedback, such as a three-dimensional (3D) computer animations of the lips and oral cavity, to illustrate

how a sound is articulated, so that a learner can use the visual model to correct their pronunciation error. Further empirical studies on this blended EFL pronunciation learning method are needed.

## References

1. Fundamentals of Artificial Intelligence, K.R. Chowdhary Springer India, 2020 . Science
2. Automatic Speech Recognition 101: How ASR systems work, Riqiang Wang, <https://www.dialpad.com/blog/automatic-speech-recognition/>
3. Text-conditioned Transformer for automatic pronunciation error detection, Zhan Zhang, Yuehai Wang, Jianyi Yang ,*Speech Communication* Volume 130, June 2021, Pages 55-63
4. Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills, Adhan Kholis *Pedagogy: Journal of English Language Teaching* Volume 9, Number 1, June 2021
5. Student Perceptions of Mobile Automated Speech Recognition for Pronunciation Study and Testing, Thomas Dillon and Donald Wells, *English Teaching*, Vol. 76, No. 4, Winter 2021, pp. 101–122
6. Improving English pronunciation via automatic speech recognition technology, Xiaobin Liu, Manfei Xu\*, Meihui Li, Meiting Han, Zejia Chen, Yiling Mo, Xiujian Chen and Minjia Liu, *Int. J. Innovation and Learning*, Vol. 25, No. 2, 2019
7. Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation, In-Seok Kim, Kim, I.-S. (2006). *Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation*. *Educational Technology & Society*, 9 (1), 322-334.
8. USING ASR TECHNOLOGY IN ENGLISH PRONUNCIATION TEACHING, Riikka Muhonen, Master's Thesis, Tampere University, Master's Programme in English Language and Literature, April 2021
9. The Role of ASR Training in EFL Pronunciation Improvement: An In-depth Look at the Impact of Treatment Length and Guided Practice on Specific Pronunciation Points, Ryan Spring, *Computer Assisted Language Learning Electronic Journal (CALL-EJ)*, 23(3), 163–185 2022
10. The Effect of Real-time Score Feedback on L2 English Learners' Pronunciation and Motivation in an ASR-based CAPT System, Reece Randall, Yeonjung Hong, Hosung Nam, *Korean Journal of Applied Linguistics* 37/4 (2021) pp.7–50
11. USING ASR TO IMPROVE TAIWANESE EFL LEARNERS' PRONUNCIATION: LEARNING OUTCOMES AND LEARNERS' PERCEPTIONS, Wen-Hsin Chen, Solène Inceoglu, Hyojung Lim, *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380–9566, Northern Arizona University, September 2019 (pp. 37–48).
12. 'Look, I can speak correctly': learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology, Muzakki Bashori, Roeland van Hout, Helmer Strik & Catia Cucchiariini, *Computer Assisted Language Learning*, <https://doi.org/10.1080/09588221.2022.2080230>
13. The Effects of Pronunciation Instruction Using Speech Recognition Software for Adult Learners of English, Tomoko Takenouchi, *The Journal of the Practical English Phonetics* 3 (2022) 1–29
14. Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their Affordances, William Gottardi, *Linguagem e Tecnologia* <https://doi.org/10.35699/1983-3652.2022.36736>

15. ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs, Solène Inceoglu, Hyojung Lim, Wen-Hsin Chen, *The Journal of Asia TEFL* <http://journal.asiatefl.org/> e-ISSN 2466–1511 © 2004 AsiaTEFL.org.
16. Assessing the Practicality of Using an Automatic Speech Recognition Tool to Teach English Pronunciation Online, Spring, Ryan1 ; Tabuchi, Ryuji2, *Journal of English Teaching through Movies and Media* Vol. 22, No. 2, (May 31, 2021)
17. Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app, Jonás Fouz-González, *Language Learning & Technology* ISSN 1094–3501 February 2020, Volume 24, Issue 1 pp. 62–85

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

