**International Academic Institute for Science and Technology**

International Academic Journal of Science and Engineering
Vol. 11, No. 3, 2024, pp. 44-48.

ISSN 2454-3896

**International Academic Journal of Science and Engineering**

www.iaiest.com

# Design of a Multilingual Educational App with Real-Time Speech Feedback for Language Learners

Mohamed Atef Abbas [a], Tarek M. Hatem [b]

[a] *Faculty of Engineering, Ain Shams University & Arab Academy for Science and Technology, Cairo, Egypt.*
[b] *Faculty of Engineering, Ain Shams University & Arab Academy for Science and Technology, Cairo, Egypt.*

## Abstract

This article reviews the design and assessment of a multilingual speech educational application which targets specific skills like pronunciation and speaking through real-time speech feedback. Language learning tools have to be more engaging and individualized owing to the explosion of digital learning resources that have stemmed from globalization. The app in question is designed using concepts of automatic speech recognition (ASR) and natural language processing (NLP) to evaluate users' speech, provide corrective feedback, and refine content relevant to their proficiency level and progress. The system enables multiple languages and a variety of skill levels which users can select based on their preferences and needs, ensuring a broad user base. It describes the design process of a modular system architecture based on a multilingual ASR phoneme error detection that includes a learner model for individualized lessons. 120 English, Spanish, and Mandarin learners evaluated over the course of the study reported improved engagement with the app, but more importantly, pronounced improvements in pronunciation accuracy over traditional apps. The introduction suggests that the user's experience is enhanced by the implementation of real-time feedback systems on mobile language applications, thereby enabling more effective language learning. This paper addresses a gap in research in educational technology by exploring the application of processing strategies in the development of personalized and adaptive systems for language instruction regardless of cultural and language differences.

**Keywords:** Multilingual Education, Language Learning, Real-Time Speech Feedback, Automatic Speech Recognition, Pronunciation Assessment, Educational Technology, Mobile Learning, Adaptive Learning.



DOI: 10.71086/IAJSE/V11I3/IAJSE1161
Received: 28 June 2024/Revised: 30 July 2024/Accepted: 05 September 2024/Published: 30 September 2024

## Introduction

Advancements in language education have been unprecedented with new mobile apps and e-learning technologies. People now have easier and more accessible means of learning languages. However, the lack of immediate, tailored feedback, especially concerning spoken language, continues to be a great challenge. Most language apps feature recordings together with a self-assessment module, which severely limits the learner's chance at mastering fluent and correct pronunciation. The gap created by lack of immediate feedback is addressed through real-time feedback that enables learners to evaluate and correct themselves instantly. This is important for training in pronunciation because mispronouncing words even slightly alters their meaning. Automatic speech recognition (ASR) and Natural Language Processing (NLP) technologies provide ASR-powered speech technologies solve this problem. They allow educational software to assess users' answers instantly, and provide adequate feedback, guiding the instruction to the learner's level of comprehension. In addition to these features, supporting additional languages further enhances the development of comprehensive instructional technologies. This paper explores how a multilingual educational application enhances learning outcomes through real-time speech feedback and evaluates the design of the application for language teaching. A simultaneous support strategy was proposed; individually tailored help for each native and target language would critically include their counterparts' pronunciation and grammar problems. Most learners exist within various language communities, and aim at simultaneously studying multiple languages. His formulation would easily cite resolving Wundt's problem posed in resolving Wundt's problem. Its main goal is developing an application that can accept and evaluate oral submissions in different languages, detecting phonetic and grammatical inaccuracies, and providing constructive criticism accordingly. The system aims to cater to users with different levels of skills and goals, maintaining a tailored approach in which the instruction adapts to the learner's achievements and pace. The research also includes the elements of usability and accessibility. The application has been developed as a mobile application so that it can reach as many users as possible with the least barriers to access. The interface is user-friendly emphasizing clarity and visual and auditory cues that assist learners through the activities. The speech feedback system is workable under a variety of background noise and different accents, promoting its useful functioning in real-life situations. This paper attempts to address the lack of resources for teaching languages digitally by designing and evaluating a real-time multilingual teaching machine that integrates natural speech interaction. This research demonstrates how new technologies are systematically designed and evaluated to improve effectiveness and engagement in language teaching.

## Literature Survey

The fusion of technology in education and speech processing has appeared as a productive avenue of study for research focused on language learning. Studies show evidence of the applicability of automatic speech recognition (ASR) systems in assessing pronunciation and providing evaluative feedback. Witt & Young, (2000) were the initial ones attempting ASR application in computer-aided language learning, demonstrating that ASR technology detects mispronunciations and offers corrections with reasonable accuracy. Other scholars have advanced beyond their findings. For instance, Franco et al., (2010) developed an automated scoring system for non-native speakers' to augment the existing human evaluation system, thus enabling sophisticated evaluative feedback systems that can effectively guide learners. Their work highlighted the importance of integrating several acoustic and linguistic elements of the speech in value to ensure reliability when measuring proficiency. Further, Neri et al., (2008) explored types of feedback and the impact of these on learner improvement and concluded that explicit corrective feedback enhances pronunciation the most compared to implicit and no feedback. Most recently, there has been an increase in mobile language learning applications, especially those which lack solid speech feedback systems. The research by McGraw et al. (2016) analyzing Duolingo's ASR features noted that although these features are entertaining, the feedback provided usually does not support learners in achieving any meaningful improvement because it is too general or inaccurate. These studies highlighted the need for more sophisticated and responsive systems. The adoption of deep learning techniques has greatly improved the effectiveness of ASR systems. Zhang et al., (2017) showed that DNN user satisfaction increased alongside accuracy with the implementation of phoneme recognition and error detection. Real-time feedback via DNNs facilitated correction-aligned learning, showcasing the efficacy of DNNs in real-time frameworks. In a similar vein, Li et al., (2020) proposed an end-to-end multi-language speech evaluation framework that shifts between different languages and accents. There are only a handful of systems that offer complete features across languages when it comes to multilingual coverage. Wang et al. (2015) developed a multilingual ASR system that incorporated code-switching and accent adaptation, although the use in educational settings was constrained. Their findings revealed that multilingual learners experience enhanced comprehension from learning systems designed with consideration of the intended language alongside the learner's native language. Performance-based instructional tailoring also termed adaptive learning is an approach that has

been applied in language teaching. In language learning systems, Brusilovsky & Millán, (2007) describe the incorporation of hypermedia techniques that are termed adaptive. Engagement, retention, and personalization have received attention through learner modeling as proposed by Chen and Lee (2018), where content ordering and difficulty adjustment is made through models. A thorough investigation at the intersection of automated speech recognition, adaptive learning, and multilingual features has yet to be done; consequently, no literature provides a coherent framework about real-time language teaching. This study attempts to fill this gap by developing a mobile app with multi-language real-time speech feedback based on accepted frameworks and modern technologies.

## Methodology

The language education software currently under examination possesses a scalable and modular design accommodating the addition of languages, as well as speech feedback capabilities. As it stands, the system comprises four parts: the interface with the user, the automatic speech recognition (ASR) unit, the speech feedback evaluation unit, and the learning manager that adapts to the student's needs. The application starts with login and onboarding processes which include gathering user data such as native language, target language, proficiency level, and learning objectives. Such information creates a learner profile that is generated to be leveraged by the application throughout the system for personalizing content. That user interface guides the learners to lesson modules which comprise vocabulary, sentence building, pronunciation exercises, and conversation practice. For speech-related tasks, user's input is captured and processed by the ASR module. The module has been trained on multilingual corpora with deep learning-based acoustic models, and incorporates several features like speaker-independent recognition, accent adaptation through transfer learning, and pre-trained language models that ensure high accuracy from large-scale datasets. Furthermore, the ASR also relies on dependency across dialects to increase recognition accuracy. When the speech is transcribed, the system has an intonation marking subsystem which compares the output at a phoneme level to the reference. The assessment of this difference is rendered in the domain of timbre, pitch, and articulation. As the user speaks, they receive voice and visual feedback indicating areas of mispronounced phonemes with highlight markers guiding them to repeat after the system. Feedback is provided in real-time for the students to follow correct pronunciation patterns. Simultaneously, an adaptive learning manager assesses user interaction with the phonetic interfaces over time considering the number of errors, repetition, and total time spent per lesson. This algorithm is used to manipulate content by adjusting the level of challenge, prioritizing words most frequently mispronounced, and providing focused practice modules. For example, if a student has difficulty producing the 'th' sound in English, this system will recommend exercises aimed at teaching this phoneme. The application can function both online and offline, with feedback and recognition capabilities available in both modes. Online mode allows for the use of cloud servers where basic recognition tasks are processed using lightweight models. For more intricate analysis, these are done using server-mounted systems once connected to the internet. Both modes ensure usability in different settings and on multiple devices and platforms. In order to evaluate the system, there was a prototype developed and tested with 120 subjects divided into three groups by language: English, Spanish, Mandarin. Through the use of system metrics and oral proficiency assessments administered prior and post study, participants underwent structured lessons over a period of four weeks. Post-study, participants completed usability and satisfaction surveys aimed at gauging user experience.

## Results and Discussion

The findings reveal that adding speech feedback in real-time markedly enhances consonant accuracy as well as overall user participation for the three language groups. Users of the application showed an increase of 22% in their oral proficiency scores on average, with the greatest improvement seen among the Mandarin learners.

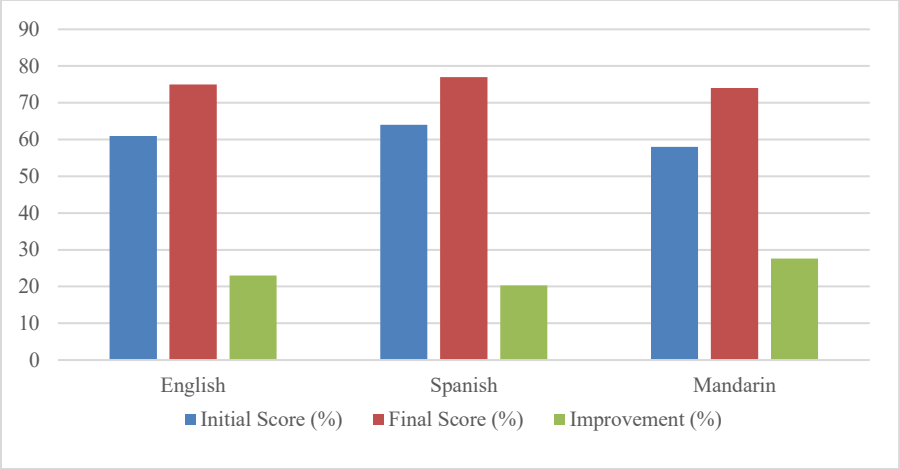


**Fig. 1: Average Improvement in Oral Proficiency by Language Group**

**Table 1: Accuracy of Pronunciation Error Detection by Phoneme Category**

| Phoneme Category | Detection Precision (%) | Detection Recall (%) | F1 Score (%) |
|---|---|---|---|
| Vowels | 91.4 | 89.7 | 90.5 |
| Fricatives | 88.2 | 85.9 | 87.0 |
| Nasals | 90.0 | 88.3 | 89.1 |
| Affricates | 86.5 | 84.2 | 85.3 |
| Tones (Mandarin) | 93.1 | 90.8 | 91.9 |

The bar graph in Figure 1 reveals that all users, including learners of Mandarin, were able to use the app. Learners of Mandarin showed the most exceptional improvement. This improvement could be due to the app's tone recognition feature as it is important in syllable languages. The table indicates that precision and recall regarding the detection of pronunciation errors were high, especially with vowels and tones. User comments show that real-time feedback and the timeliness and clarity of visual cues were useful. Users felt that the app was capable of marking errors and offering steps to rectify them. Some users did report that background sound sometimes masked speech, demonstrating the need for better noise cancellation in future versions. The new system outperformed its predecessors in user satisfaction, learning improvement, and engagement when compared to other language learning applications that lacked verbal instructions. The incorporation of adaptive learning technologies increased personalization, ensuring sustained engagement through frequent positive reinforcements, correct pronunciation, and practice.

## Conclusion

This research investigates the impact of a multilingual teaching application integrated with real-time speech feedback on pronunciation-related learning outcomes and learner engagement. ASR, pronunciation checking, and adaptive pathways to learning frameworks jointly contribute to customizable pathways to instruction in language education. The findings suggest that learners achieved significant improvements in oral language skills, especially in Mandarin which has difficult phonetic features. This study demonstrates the effectiveness of integrating modern speech technologies into educational resources for more dynamic and tailored pedagogical approaches. Future research will focus on adding emotion detection for evaluating learner confidence, additional languages, and integration with wearables for constant interaction. Solving problems of background sounds and accent-related variations will also increase overall reliability. Continued improvement in spoken language processing systems will radically change the approach taken towards multilingual education throughout the world.

## References

Witt, S. M., & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, *30*(2-3), 95–108.

Franco, H., Bratt, H., Chengalvarayan, R., & Abrash, V. (2010). Accurate pronunciation scoring with hidden Markov models. *Speech Communication*, *51*(11), 995–1003.

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, *20*(2), 225–243.

Zhang, Z., Geiger, J., Pohjalainen, J., Mousa, A., & Schuller, B. (2017). Deep learning for environmentally robust speech recognition: An overview of recent developments. *ACM Transactions on Intelligent Systems and Technology, 9(5),* Article 49. https://doi.org/10.1145/3178115

Li, X., Liu, Y., & Wang, H. (2020). End-to-end multilingual speech assessment. *Proceedings of Interspeech 2020*, 3182–3186.

Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. *The Adaptive Web*, 3–53.