IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 April 2024, accepted 25 May 2024, date of publication 28 May 2024, date of current version 10 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3406494

RESEARCH ARTICLE

Interactive Augmented Reality System for Learning Phonetics Using Artificial Intelligence

RAHMA M. TOLBA [1], TAHA ELARIF [1], ZAKI TAHA [1], AND RAMY HAMMADY [2], (Member, IEEE)
[1]Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
[2]School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K.

Corresponding author: Rahma M. Tolba (rahma.tolba@cis.asu.edu.eg)

This work was supported by the School of Computer Science and Electronic Engineering, University of Essex, through the Open Access Fund.

ABSTRACT The increasing adoption of language learning apps that utilize Augmented Reality (AR) and Artificial Intelligence (AI) for speech recognition has sparked interest in the potential benefits for phonetics education. However, currently available AR apps only focus on teaching letter names and vocabulary, lacking the potential for a more immersive learning experience. To address this limitation, this paper introduces an interactive AR system that integrates AI speech recognition with AR to provide an engaging and interactive learning experience. To showcase the capabilities of the proposed system, we have created a prototype for the Arabic Phonetic Atlas textbook. This prototype enhances reading the /s/ sound page in the Atlas by incorporating a 3D animated model of the speech organs onto the existing 2D image. The dynamic animation of the 3D model reflects the sound description provided in the Atlas. The system also offers real-time user pronunciation feedback through a customized AI phoneme recognition system. A comprehensive user study was conducted to evaluate the usability and learning impact of the proposed system, involving 83 adult participants aged between 18-40. The assessment approach involved the use of both direct and indirect observations, as well as various surveys to gather both numerical and qualitative information. The findings indicate not only a greater level of understanding compared to conventional methods but also an improved capability to master specific phonemes quickly and effortlessly. In addition, they are showing great potential for the proposed system to be incorporated into the conventional classroom setting as an instructional aid.

INDEX TERMS Interactive augmented reality, learning phonetics, language learning, Arabic phonetics.

I. INTRODUCTION

Learning phonetics (LP) is a key component in learning a language, as it involves the analysis of the sounds that form a language and their arrangement. Knowledge of phonetics is crucial in accurately producing and understanding the sounds of a language, which is vital for effective communication [1]. Moreover, phonetics plays a significant role in developing literacy skills and aiding in decoding and reading fluency. LP facilitates language acquisition by enabling learners to recognize and imitate the subtle variations in pronunciation. Therefore, having a solid grasp of phonetics is indispensable for achieving success in language acquisition [2]. Computer Assisted Pronunciation Training (CAPT) is an innovative approach to language learning that utilizes technology to improve students' proficiency in producing clear and accurate speech. It offers an interactive and comprehensive system for practicing pronunciation [3]. By combining modern technology like Augmented Reality (AR) and Artificial Intelligence (AI) with traditional language teaching methods, CAPT provides instant feedback on articulation through specialized software, which helps learners identify and rectify pronunciation mistakes.

The utilization of AI in language learning apps has been on the rise, thanks to the implementation of machine learning algorithms that analyze user progress and offer personalized

The associate editor coordinating the review of this manuscript and approving it for publication was Ganesh Naik.

feedback and suggestions. In 2015, ELSA Speak platform received recognition from Forbes for its use of AI to bring about positive change. It has also been ranked in the Top 5 AI Apps, alongside major players like Microsoft's Cortana and Google's Google Allo. Its flagship product, Speech Analyzer, uses AI technology to serve as a conversational coach for English fluency, providing instant feedback to users. With its impressive capabilities, ELSA Speech Analyzer empowers individuals from diverse backgrounds to communicate confidently in English [4]. The study conducted by Zou et al. [5] aimed to determine the effectiveness of utilizing automatic feedback from AI speech evaluation programs in enhancing the speaking proficiency of EFL (English as a Foreign Language) learners. In Dalim et al. study [6], the effectiveness of a novel teaching strategy combining AR and speech recognition technologies was explored in terms of knowledge gain and enjoyment for children learning English vocabulary for color, shapes, and spatial relationships.

With the integration of AR technology, learners can experience an interactive and immersive environment that facilitates the acquisition of phonetic skills. It allows the learners to visualize articulatory movements for the accurate production and perception of speech sounds. There are various types of AR applications, such as mobile-based and computer-based applications [7]. Using mobile devices as a platform for AR has introduced a new level of accessibility and convenience for users. By utilizing the advanced capabilities of mobile devices, such as cameras and sensors, AR technology can create a realistic and interactive experience for users [8]. This has led to the adopting of mobile-based AR (MAR) in various fields, including education, entertainment, and marketing. Our previously published study focused on MAR applications used for LP [9]. MAR for LP has been explored in many languages, such as English [10], [11], Japanese [12], and Chinese [13]. Nugraha et al. [14] provided an illustrative case of integrating AR in the language classroom as an effective pedagogical tool for teaching English phonetics to young learners. In the study by Aladin et al. [15], an interactive AR tool, "AR-To-KID," was developed for preschool children aged five to six years. The goal was to examine the potential of AR in enhancing English pronunciation learning among young learners, specifically through incorporating speech input in the AR interface. Afriliani et al. [16] developed an English phonetic model for literature students to facilitate self-directed learning via MAR.

Up to the point of writing this paper, there is limited research on the use of MAR technology for learning phonetics of the Arabic language. Yet, no specialized apps were found that specifically targeted this skill. However, a recent study by Daud et al. [17] introduced a MAR application named ARabic-Kafa, which focuses on developing educational materials for Arabic vocabulary learning. Phonetics was not the main focus of their study. In terms of visualization, the majority of MAR applications available offer users access to 3D representations of Arabic letters, accompanied by audio for the proper pronunciation, such as Mondly AR [18] and Arabic [19]. None of the MAR apps currently offer a 3D guided animation showcasing speech organs' movement. Only one PC program, "Quran-Tajweed" [20], was found to offer a 3D model of the jaw, teeth, and tongue for learning Arabic phonetics. However, when it comes to learning Arabic phonetics, a more comprehensive model that includes other speech organs, such as the vocal tract, is needed. This is because Arabic contains two letters and four sounds pronounced from the throat [21].

Learning phonetics is challenging for students for several reasons, such as the complexity of the phonetic system, the limited availability of high-quality learning resources, and the presence of unfamiliar sounds [22]. The failure to use and understand phonetics leads to mispronunciation issues, which, in turn, impede both oral and written communication. Current AR Arabic language learning systems mainly focus on the Arabic alphabet and vocabulary pronunciation. However, our research introduces an innovative AI MAR system that provides a dynamic and interactive learning experience. To achieve this, we utilize 3D animated models that visually guide learners in the precise positioning of the tongue and other speech organs necessary for accurate phoneme pronunciation. This immersive approach allows learners not only to hear but also to see the intricate details of proper phoneme formation.

Furthermore, our system incorporates interactive student pronunciation feedback using AI automatics speech recognition (AI ASR) system. After attempting to pronounce a phoneme, students receive immediate feedback based on their pronunciation. Correct pronunciation unlocks the opportunity to practice words, enabling learners to reinforce their newly acquired phonetic skills. On the other hand, incorrect pronunciation activates guided animations that are specifically designed to address and rectify the issues causing inaccurate pronunciation. This feedback mechanism ensures that learners receive personalized support in mastering phonetics pronunciation, just like they would receive from a language teacher in the classroom. Figure 1 illustrates the enhanced interactivity features that our research brings to the field of MAR-based language learning (LL) applications.

We perform an extensive user study to evaluate the proposed system's usability, effectiveness, and learning impact. The participants of the study were chosen based on their diverse backgrounds in the Arabic language. The interactivity of the system was also thoroughly analyzed through a group-based approach. The collection of quantitative data from both controlled and experimental groups enabled us to thoroughly evaluate the usability of AR technology in the educational process. Furthermore, the effectiveness and impact of the interactive AR educational environment were extensively analyzed, which is crucial in determining the suitability of remote learning tools in an educational context [23]. Hence, the comprehensive analysis in this study, combined with a user study comparing AI AR-based learning with traditional

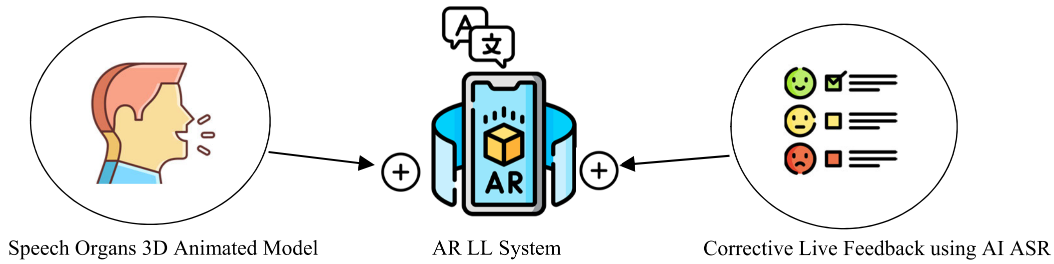


FIGURE 1. The enhanced interactive and innovative features that our research brings to the field of MAR-based language learning applications.

methods, strongly suggests that interactive AR-based learning offers a promising alternative to traditional face-to-face learning in exceptional circumstances. To the best of our knowledge, this is an entirely novel system, and no other AR system currently exists that utilizes AI for phonetics learning in general or for Arabic phonetics specifically.

II. SYSTEM ARCHITECTURE

The pipeline structure consisted of three key phases, as depicted in Figure 2. First, meticulous preparation was carried out for the audio and visual content. A careful data collection method was employed to generate the phoneme audio dataset, followed by a thorough screening process to eliminate imperfect or substandard recordings. Then, the filtered recordings were meticulously inspected by a target language's phonologist, who provided appropriate labelling to differentiate between correct and incorrect pronunciations. Creating the 3D animated model involved capturing dynamic MRI images of a phonologist articulating the target phoneme, which was later utilized for animating the 3D model. The MRI recordings, along with phoneme descriptions from the language phonologist, were provided to a skilled 3D modeler and animator to create the final 3D animated model. Moving onto Phase II, the focus shifted towards developing and training an AI ASR system that would provide students with feedback on their pronunciation. An API was then created to facilitate easy updates to the trained model. Finally, in Phase III, the AR application was built, encompassing a user-oriented UX/UI design approach. A suitable game engine and AR software development kit (SDK) were carefully selected to cater to the needs of the AR app. Extensive testing was conducted, followed by necessary maintenance before the app's official release.

The proposed system includes specific features chosen carefully to fulfil the interactivity aspects of the learning process. All proposed system features are listed in Table 1. The features were discussed with the language phonologists and teachers to ensure its optimization. The 3D animated model of phoneme pronunciation is a cutting-edge educational tool that combines medical imaging technology with linguistics, illustrating how phonemes are produced with precision. The unique aspect of this model lies in its incorporation of MRI recordings that show the anatomy and dynamic motion of the vocal tract during speech production in detail. In order to offer precise and accurate phoneme pronunciation, the correct pronunciation of each phoneme was recorded by phonologists of the target language. A customized AI phoneme recognition system is utilized to check students' pronunciation. This system is specifically trained to recognize correct and incorrect pronunciations of each phoneme to provide live, immediate feedback to the student, similar to what language teachers do in classrooms. The proposed system offers a solution to incorrect pronunciation by providing a guided 3D animated model to demonstrate the correct pronunciation of the target phoneme. This system includes multiple animations targeting common issues identified by a speech therapist. Additionally, carefully selected practice words are incorporated in collaboration with the target language teacher to reinforce the correct pronunciation learned by the student.

III. SYSTEM DEMO

The proposed system was utilized to learn the 'Sīn /s/' sound in Arabic. Learning the Arabic language is considered difficult due to the presence of another letter, 'Ṣād /sˤ/', with a similar articulation to 'Sīn /s/'. This letter is unique to Arabic and poses a challenge for both native and non-native speakers. Collecting audio data revealed that 70% of the recordings for 'Sīn /s/' were pronounced inaccurately, even surprising the phonologist. These incorrect pronunciations were not due to physical limitations but rather a lack of understanding about the correct pronunciation of the letter. As step 1 in Figure 2 illustrates, our first task was to gather all the necessary materials that were previously missing. First, we recorded the proper pronunciation of /s/ and /sˤ/ with the help of an Arabic phonologist. The audio recordings collected were then filtered to produce the audio dataset needed for training the AI ASR system. This system was developed specifically for this project using LSTM and Python to take in the student's pronunciation and provide immediate corrective feedback based on it [24]. Next, we recorded the Arabic phonologists pronouncing /s/ and /sˤ/ through dynamic MRI scanning, as depicted in Figure 3. This enabled us to clearly distinguish between the two sounds and serve as a reference for the subsequent creation of the 3D model animations of speech organs. We then developed a 3D model (Figure 4) that met our criteria of being informative and approachable for

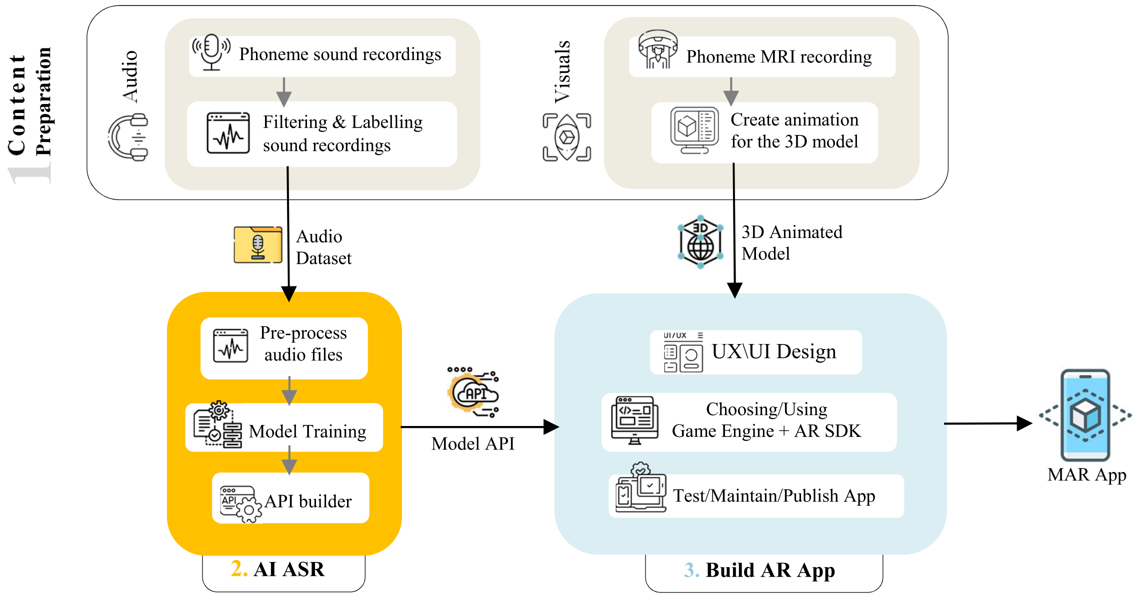


FIGURE 2. The proposed system architecture.

TABLE 1. List of proposed system features.

Feature	Description	Purpose
3D Animated Model	A 3D animated model of the articulation points of each phoneme. It is based on real-time MRI recordings for language phonologists, along with detailed phoneme descriptions provided by the phonologist.	It provides a comprehensive understanding of how phonemes are produced by the human speech organs, unlike the existing 2D images that are used.
Audio	The correct pronunciation of each phoneme. It is recorded by the phonologist of the target language.	Ensure that every student has access to the correct and accurate phoneme pronunciation.
Pronunciation Check	A customized AI speech recognition system trained on the correct and wrong pronunciation of the target language phoneme.	It offers live feedback to the student, mimicking the language teacher's role. This feedback mechanism provides a supportive and engaging environment for the student, ensuring their progress and success in language acquisition.
Guided Animation	Multiple guided animations for the common reasons for incorrect pronunciation of each phoneme.	Mimicking the role of the language teacher and speech therapist by showing the students which part of the speech organs they need to focus on to pronounce correctly.
Practice Words	Carefully selected words that contain the target phoneme. They are chosen with the help of the target language teacher.	Allowing learners to practice and refine their ability to produce that phoneme accurately.

students. With the help of an Arabic language speech therapist, we compiled a list of all possible reasons for students' incorrect pronunciation of the letter ''Sīn /s/''. Based on this, we created four animations specifically for the /s/ sound. One animation demonstrated the proper pronunciation with an MRI recording as a reference, while the other three provided step-by-step guidance for correcting incorrect pronunciation in real time. The sequence of images for the animation of the correct pronunciation and the other three cases can be seen in Figure 5 and Figure 6, respectively. Finally, we selected the practice words from the Arabic phonetic Atlas (Table 2).

Figure 7 illustrates the process of the student interacting with the AI MAR application. With the smartphone camera directed towards the target image of the /s/ phoneme (from the Arabic Phonetic Atlas), the MAR application immediately displays a 3D model and the phoneme's correct pronunciation. The interface also includes options to replay the animation and listen to the pronunciation again. Furthermore,

TABLE 2. **The used practice words.**

Sīn /s/
سلام
الإسلام
حارس

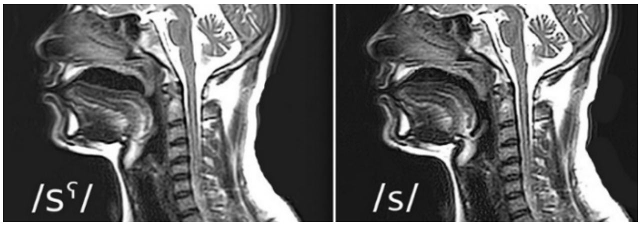


FIGURE 3. **MRI images for articulation point of** "Sīn /s/" **on the left and** /sˤ/ (ṣād) **on the right [9].**

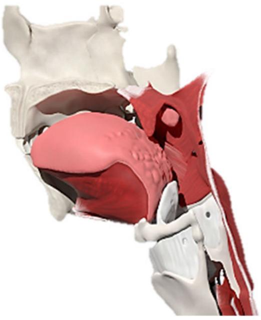


FIGURE 4. **Screenshot of the created 3D model.**

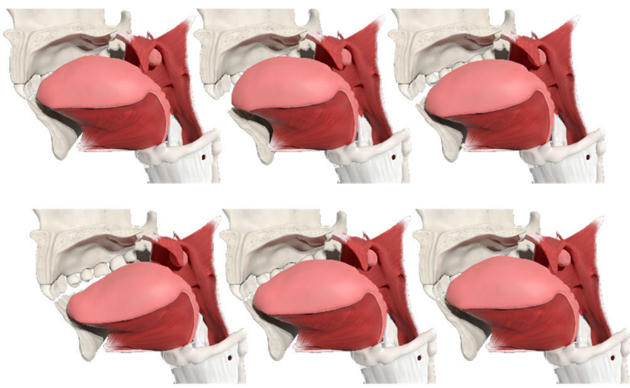


FIGURE 5. **A sequence of images for the animation of the correct pronunciation of the/s/ phoneme.**

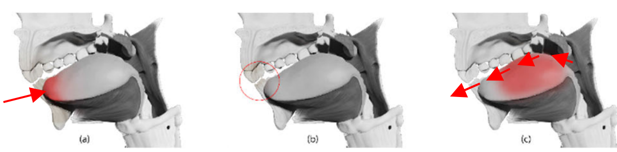


FIGURE 6. **All possible reasons for incorrect pronunciation of the letter** /s/ (sīn) **(a) Tongue tip position, (b) space between upper and lower jaw (not fully closing the teeth), (c) Tongue body shape and Airflow.**

a microphone button is provided to encourage the student to practice pronouncing the phoneme independently. The application records the student's pronunciation and gives instant feedback based on their performance. If the pronunciation is accurate, the student will be given practice words. However, if there are mistakes, the system will guide the student through a series of animated visuals to help them correct their pronunciation.

The developed application encompasses various menus and options, including choosing the display language and accessing App tutorials. The App Tutorial is a comprehensive guide for navigating and utilizing the application. The screenshots in Figure 8 showcase the app's different features, such as the display language menu, welcome message, main menu, list of Arabic letters, and the AR scene. Before start learning, students have the option to choose the language they prefer to see on the display. Furthermore, they can also access a tutorial within the app that gives them a quick explanation of its features. The "Choose a Letter" option lets students view a menu containing all the Arabic letters. They can then select the specific letter they wish to study. Once selected, the mobile device's camera is activated, allowing students to scan the page devoted to the /s/ sound in the Arabic phonetics Atlas. This triggers the appearance of a 3D animated model of the speech organs, accompanied by an audio demonstration of the correct pronunciation of the /s/ sound. Additionally, control buttons are available to play the 3D model, return to the main menu, or begin practicing pronouncing the /s/ sound and receive instant feedback.

IV. SYSTEM EVALUATION

Phonetics is a fundamental aspect of language acquisition, and its mastery plays a crucial role in the overall development of language skills. Therefore, it is essential to assess the phonetics learning process to understand the effectiveness of the proposed system as a learning tool and its impact on the learning process. The effectiveness was measured in terms of usability, user experience (UX), and interactivity [23]. Assessing the learning impact can be effectively accomplished through various methods that target different aspects of phonetic knowledge and skills. Phonetic discrimination tasks evaluate students' capacity to differentiate between similar phonemes, which is crucial for developing auditory discrimination skills [25]. Speech production tasks, whether self-assessed or instructor-evaluated, gauge pronunciation accuracy and clarity. While pre- and post-assessments measure learning growth over time [26]. To evaluate the proposed system, we conducted a quasi-experimental research design with two groups: an experimental group and a control group [27]. Figure 9 outlines the steps involved in the experiment. The production tests (pre, post, and delayed) completed by the participants were first analyzed quantitatively. Then, the qualitative data from surveys were examined to complement and clarify the quantitative results. Eighty-three participants, who were native Arabic speakers, voluntarily participated in the study. They were randomly divided into the experimental group ($n = 43$, 22 males and 21 females) and the control group ($n = 40$, 20 males and 20 females). The participants were between 18 and 40 years old and had

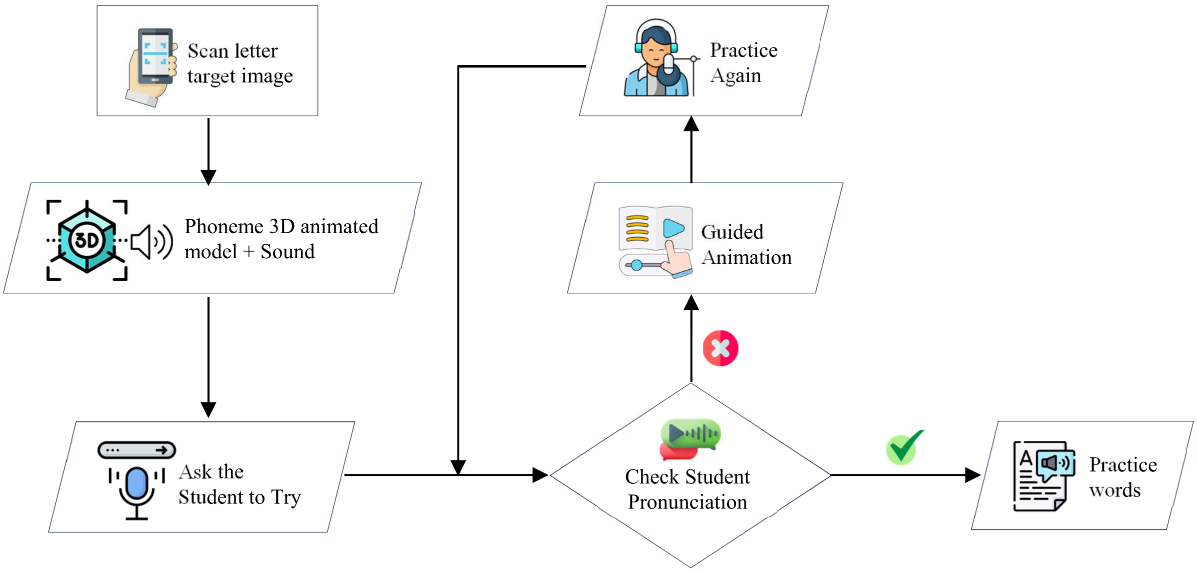


FIGURE 7. Interaction flowchart that occurs between the student and the AI MAR application.

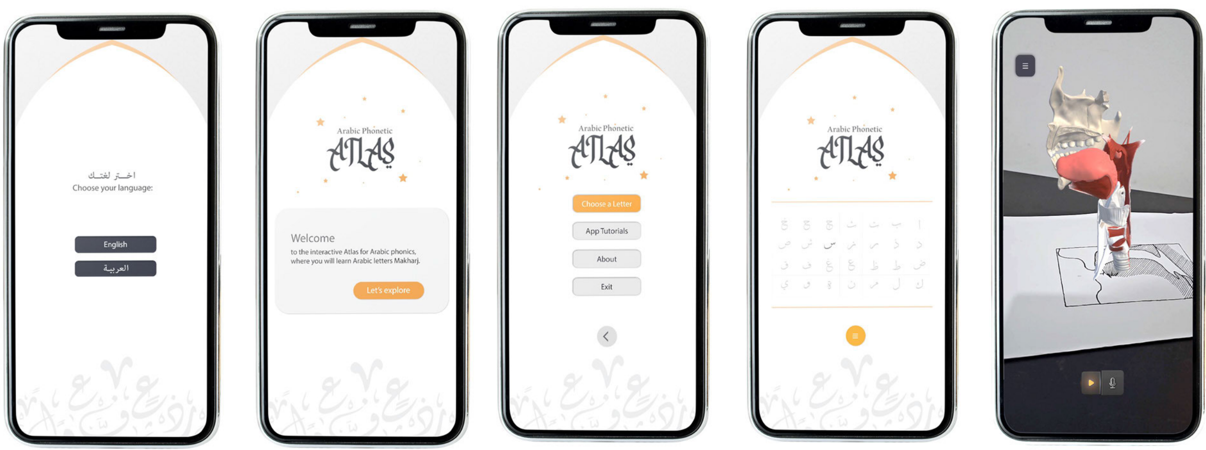


FIGURE 8. Screenshots of the developed MAR application.

intermediate proficiency in Arabic. None of them had previously taken any phonetics courses.

A. ARTICULATORY AWARENESS QUIZ

The articulatory awareness quiz was completed by all participants twice: first at the beginning (see Appendix A) to evaluate their understanding of Arabic phonetics, and second after the Delayed test (see Appendix B) to measure the effectiveness of the learning session and the level of comprehension of the taught phoneme [28]. Participants were instructed on the articulatory gestures involved in producing the/s/ sound. They were then asked questions such as, "When you articulate /s/ in 'سما', do you raise your tongue high in the mouth until it almost touches the alveolar ridge?" Additionally, we adapted the phoneme-video pairing test, where participants were shown short videos of movements for a particular phoneme. An example of one of the questions is shown in Figure 10: Does the video correctly demonstrate the tongue movement required for the /s/ sound in 'sin'? (Yes/No).

B. LEARNING SESSION

The learning sessions will focus on teaching the specific phoneme sound and providing additional phonetic practice afterwards. After conducting the pre-test, the participants received an introduction to Arabic phonetics. Next, both groups were taught the /s/ sound. The experimental group was trained using the AI MAR app, while the control group used the traditional methods of learning phonetics. Both groups' teaching materials and stimuli were sourced from the Arabic Phonetic Atlas textbook, which contains comprehensive phonetic knowledge of Arabic phonetics. Finally, the participants

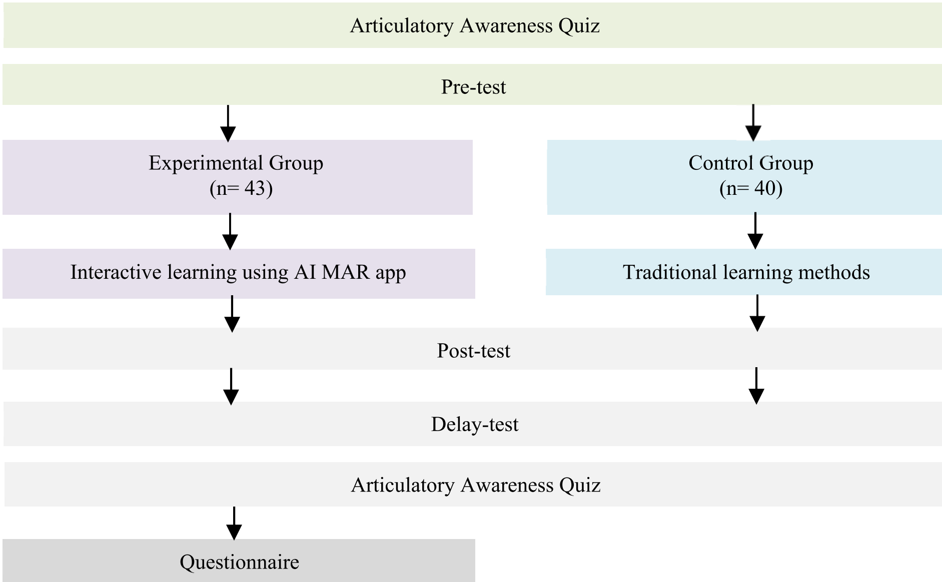


FIGURE 9. **Quasi-experiment procedure [27].**

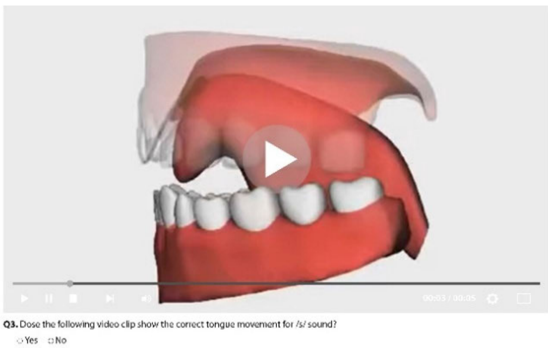


FIGURE 10. **Screenshot of a question in the articulatory awareness quiz.**

were afforded time to engage in training and practice. The training material for the participants encompassed approximately 20 words, representing all possible phonological environments for the/s/ sound. Quantitative data, including learning duration and the number of practice sessions, were recorded for both groups.

C. PRODUCTION TESTS

We conducted three production tests, pre-test, post-test, and delay-test (Appendix C). All the participants had been asked to take the pre-test before the instruction. After completing the learning session, all participants were given a post-test to measure their improvements in pronouncing the target phoneme. After three days, all participants had a delayed test to measure the retention of learned knowledge. The tests were a read-aloud task, including nine words for the target phoneme. The tests covered all the possible phonological environments and syllable positions of the phoneme taught during the sessions. The participants were given 3 seconds to read each word on the screen [28]. The data collection included audio recordings of each participant's reading performances during production tests.

D. QUESTIONNAIRES

Finally, the students completed a 16-question questionnaire with Likert scale ratings (Appendix D) to evaluate the system as a learning tool, its impact on learning, and interactivity level. Students also provided open-ended feedback about the UX and other issues, which was used for the qualitative analysis.

V. RESULTS AND DISCUSSION

To assess the statistical significance of our findings, we formulated null hypotheses positing that the language proficiency level of students has no bearing on their duration of study. In the interest of consistency, we opted for a one-way analysis of variance (ANOVA) for all production tests. It calculates an F-statistic, which is a ratio of two variances. The p-value associated with the F-statistic in ANOVA measures the likelihood of observing the calculated F-statistic (or more extreme values) if the null hypothesis were true. A small p-value ($<\alpha$, often 0.05) suggests significant differences between group means, leading to rejecting the null hypothesis. Conversely, a large p-value suggests insufficient evidence to reject the null hypothesis [29]. In this study, a significance level of alpha $= 0.05$ was chosen for ANOVA, and the null hypothesis was rejected if the p-value was less than alpha.

A. PRE-TEST RESULTS

The results (Table 3) indicate no significant impact between the two groups, as evidenced by a low F-value of 0.97 and a corresponding p-value of 0.32. Therefore, the null hypothesis, which posits that the proficiency level of both groups doesn't influence pre-test scores, cannot be rejected based on these findings.

B. POST-TEST RESULTS

The outcomes of the post-test phase reveal a noteworthy advantage of the experimental group over the control group (Table 4), highlighting the significantly improved learning outcomes facilitated by the AI MAR application in contrast to conventional methods. The increase in the F-value to 1.37 in the post-test suggests that the gap in test scores between the two groups has been reduced due to interventions such as teaching or practice.

C. DELAYED TEST RESULTS

Table 5 highlights a substantial difference between the two groups regarding retained knowledge, as indicated by a high F-value of 3.72, coupled with a low p-value of 0.06, providing strong evidence against the null hypothesis. Therefore, the null hypothesis cannot be rejected, suggesting that the test scores do not vary significantly across different language proficiency levels based on the analysis conducted. The higher F-value (3.72) indicates that the impact of language proficiency on test scores becomes more apparent over time.

D. QUESTIONNAIRES RESULTS

The results of the questionnaire regarding the usability and impact on learning of the AI MAR app were highly positive, as shown in Figure 11. It was noted that the participants had limited experience with Arabic phonetics, with only 14% reporting no experience and 28% reporting a rating of two. None of the participants rated themselves as highly experienced. Most participants (58%) strongly agreed with the statement that the AI MAR app effectively taught them the procedure of pronouncing the /s/ sound. When asked if the AI MAR app was more effective than the printed version, 86% of participants strongly agreed with this statement. While some responses may be influenced by the novelty factor associated with AR, overall, the participants found the app compelling and interesting.

Additionally, most students had a highly positive attitude towards the app and its effects on their learning, except those who responded to questions Q#9 and Q#7. Furthermore, 86% of students reported that the AI MAR app provided a meaningful and valuable learning experience in Q#10. However, some students held different opinions when it came to the possibility of using the app as a complete replacement for traditional face-to-face learning. According to the responses received from the qualitative questionnaire, most students do not see the app as a substitute for real-world learning, which is reflected in their answers to question Q#11. Additionally, the majority (85%) of students prefer to use the app as a supplementary tool to complement traditional face-to-face learning, as shown in their answers to question Q#12. This is because the interactive nature of the app allows students to practice at their own pace, which is especially beneficial when they have limited practice time in a face-to-face setting. To further assess the effectiveness of interaction while using the application, a survey was administered with four questions rated on a 5-level Likert scale. Most responses were highly positive, as depicted in Figure 12. All the participants agreed that the AI MAR app encouraged interaction and that 60% would keep using it. 84% of participants found the live feedback feature of the AI MAR app to be highly helpful, allowing participants to learn at their own pace.

The results of the pre-test indicate that the majority (87%) of the participants mispronounce the /s/ sound in Arabic when it is accompanied by a heavy sound (known as Tfkhem), wherein the back of the tongue is elevated, as observed in words 'وسط' and 'جرس' scores in Figure 12. The same was observed in the post-test words 'قوس' and 'درس', as illustrated in Figure 13. This discrepancy emphasizes the significance of the "Practice Words" step in our proposed system. It plays a crucial role in helping students master the /s/ sound, offering targeted repetition and exposure to various contexts necessary for efficient learning. By focusing on words containing the /s/ sound, learners engage in structured practice, reinforcing accurate sound production and integrating language skills. This step boosts confidence through successful outcomes and confirms the ability to utilize the acquired skill in practical communication scenarios.

The results from the articulatory awareness test (Figure 14) demonstrate the crucial role of explaining the functioning of the speech organs in producing phoneme sounds. This step is particularly significant in the process of phonetic learning, unlike the traditional approach adopted in the works of [17], [19], and [30], where emphasis is placed on teaching the letters' names, visual representation, and pronunciation within words. Learners can mimic and produce accurate phoneme sounds more effectively by knowing which speech organs are involved and how they move. Overall, it enhances learners' phonemic awareness, pronunciation accuracy, and linguistic competence. The comparison between the results from Q1 and Q3 revealed an increase in student comprehension, with scores rising from 35% to 60%. This can be attributed to the 3D animated visualization of the speech organs in Q3, which significantly improved from the written format used in Q1. This finding is consistent with the Afriliani et al. study [16], which utilized AR technology to present a 3D simulation of speech organs for learning English phonetics. It further supports the notion that incorporating 3D animated models can enhance the learning process and improve student comprehension, as demonstrated in our study.

The study's findings validate the hypothesis that integrating AI and AR in phonetics learning significantly improves the learning outcome compared to traditional approaches. Table 6 compares our prototype and two traditional learning methods (Arabic Phonetics Atlas and learning videos). The comparison reveals several advantages of our prototype. While the traditional Atlas offers 2D illustrations of the sagittal view of speech organs in black and white, our prototype presents 3D models that can be explored from

TABLE 3. The single-factor ANOVA analysis for the pre-test scores.

Test	Groups	N	Mean	SD	F (1,80)	P-value
Pre-test	Experimental	43	7.51	0.83	0.97	0.32
	Controlled	40	7.35	0.74		

TABLE 4. The single-factor ANOVA analysis for the post-test scores.

Test	Groups	N	Mean	SD	F (1,80)	P-value
Pre-test	Experimental	43	7.88	0.88	1.37	0.24
	Controlled	40	7.65	0.83		

TABLE 5. The single-factor ANOVA analysis for the delayed test scores.

Test	Groups	N	Mean	SD	F (1,80)	P-value
Pre-test	Experimental	43	8.26	0.82	3.72	0.06
	Controlled	40	7.88	0.79		

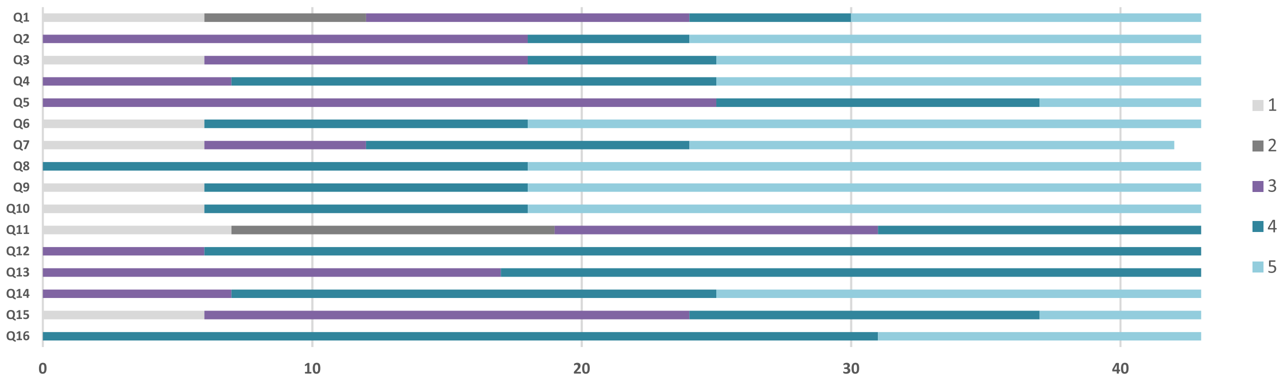


FIGURE 11. The usability and impact on learning questionnaire responses.

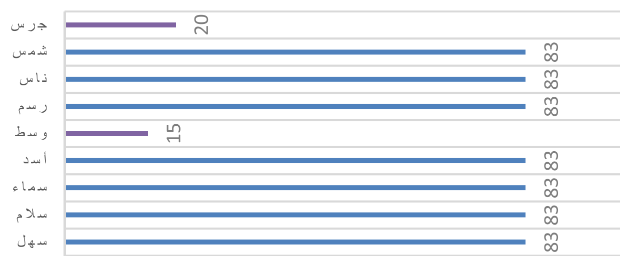


FIGURE 12. Pre-test scores.

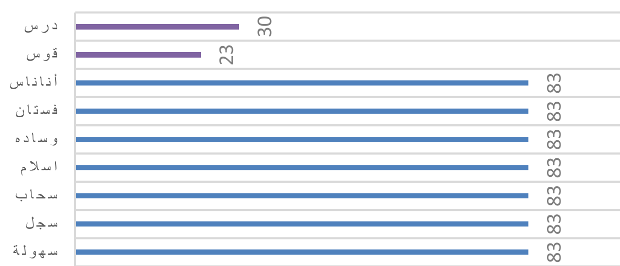


FIGURE 13. Post-test scores.

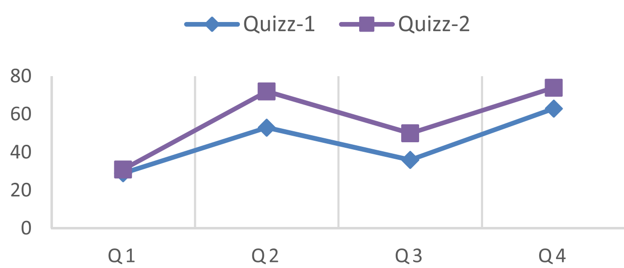


FIGURE 14. Results of articulatory awareness quizzes.

various angles, providing better understanding. Whereas learning videos only show pre-recorded sounds with static images of speech organs, or 2D animation or pre-rendered 3D animation that students can only watch without being able to interact with.

Moreover, our prototype incorporates animated demonstrations of phoneme sounds, more comprehensible than classical Arabic words in the traditional Atlas or voice recording of language teacher explanations in the learning videos. The traditional Atlas relies on ultrasound images as a guide, which only provides a horizontal view, limiting the information on the movements of the speech organs. In contrast, we utilized dynamic MRI recordings, which offer a more

TABLE 6. Traditional arabic phonetics atlas vs our prototype.

Feature	Proposed System Prototype	Traditional Atlas	Learning Video
Phoneme Visualization	3D Lively animated model	Static B&W 2D illustration	Phoneme sound with static image or 2D animation or pre-rendered 3D animation or 3D printed model of mouth (jaws, tongue, and teeth only)
Phoneme Sound Description	Dynamic: An animation.	Static: Just text adjectives written in the classical Arabic language	Static: pre-recorded voice explanation in one language
Speech Organs Movements Reference	Dynamic MRI recordings	Ultrasound images	Language Teacher / Therapist
Interactivity	• Interaction with the 3D animated model through scaling and rotation. • Voice recording option with Automatics live feedback of student's pronunciation	Non	None
Accessibility	Through any computer or Android mobile device.	B&W printed version or PDF.	YouTube or a website page
Display Language	Available in Arabic and English	In Arabic only	In one language only

comprehensive view of the speech organ movements. In terms of interactivity, the traditional learning methods lack any interactive features, while our prototype allows students to interact with the 3D models and receive real-time feedback on their pronunciation through AI technology. Lastly, our prototype is easily accessible through any computer or Android mobile device. In contrast, the traditional Atlas is only available in the form of a black-and-white printed version or a PDF.

Up to the point of writing this paper, no AR system currently exists that utilizes AI for phonetics learning or Arabic phonetics specifically. Many studies, such as [17] and [30], have focused on teaching Arabic vocabulary, but not specifically phonetics. While other studies [10], [11], [12], [13] have incorporated audio for phoneme sound, none have included a recording option using AI to analyze student recordings and provide feedback on pronunciation. Regarding incorporating a 3D animated model, the only similar study we found was that of Afriliani et al. [16] for English phonetics. In addition to Quran-Tajweed [20], a PC program for learning Arabic phonetics not a MAR app. Figure 15 compares the 3D models utilized in these studies and the one we proposed. It was observed that the model proposed by Afriliani et al. [16] was not entirely 3D but rather a 3D image of speech organs. In contrast, the 3D model option in Quran-Tajweed [20] lacks the inclusion of necessary speech organs such as the throat, which is crucial for learning six sounds in the Arabic language [21].

VI. LIMITATION AND FUTURE RESEARCH

In the open-ended feedback, it was discovered that participants anticipate having audio and video explanation capabilities in addition to the existing features. However, it was decided not to implement a video communication feature since it would seem out of place in an AR system. Instead, the lesson was designed to mimic a traditional classroom setting where the student can both see and hear the teacher's explanation and repeat it. However, it was noted that users expected a simpler explanation using the 3D animated model for a mobile app. This is a valid critique and should be considered in future work.

Moreover, additional data logging should be implemented to obtain more quantitative data on the app's usage and improve user study. This would provide a more precise measure of the ease-of-use factor, particularly regarding error rates. Q#11 in the survey also revealed that participants do not see the app as a complete replacement for face-to-face learning. This is unsurprising, as most students preferred face-to-face learning, with the app as a complementary tool.

Nonetheless, the usage data and survey results clearly show that the app was highly interactive and effective for practicing at home, even without the explanation option.

Furthermore, we received very positive feedback regarding the system's UX. The app's overall experience was highlighted as positive for its lack of complexity. The negative feedback was the unusual places for the exit and back buttons.

This study only tested the/s/ sound in Arabic. In the future, our goal is to include all the sounds of the Arabic language and develop a complete version of the AR Arabic Phonetics Atlas. We also aspire to expand the scope of the proposed system to include other languages with complex phonetic sounds, such as German and Chinese. The proposed system offers numerous benefits, primarily its ability to facilitate self-learning and practice in the comfort of one's own home, leading to improved performance in language learning. The acquired knowledge can potentially advance speech therapy, provide a comprehensive understanding of the anatomical aspects of speech for medical students, and explore the intricate phonological patterns present in various languages. Furthermore, it has the potential to aid in the accurate treatment of speech and language disorders (e.g. lisp correction).

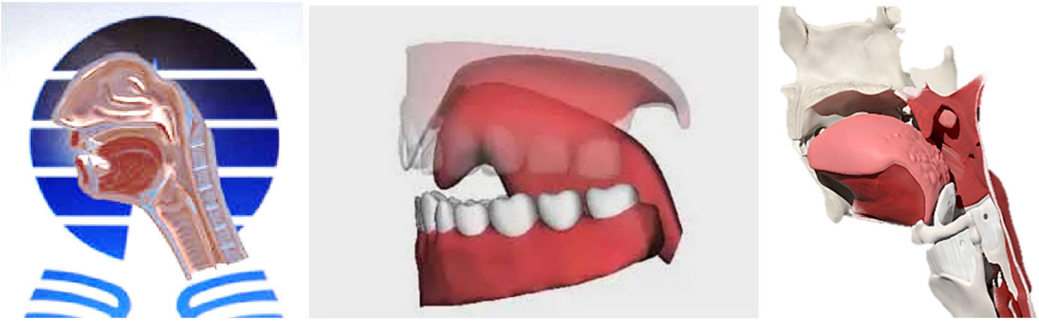
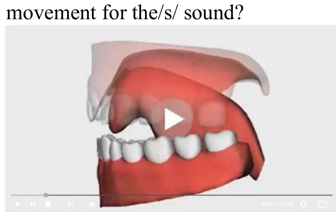


FIGURE 15. A comparison between the 3D models utilized in the Afriliani et al. study [16] on the left, the 3D model option in Quran-Tajweed [20] in the middle and the one we proposed on the right.

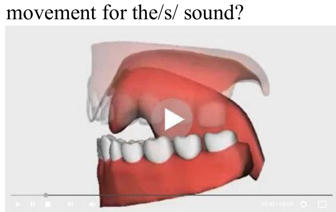
VII. CONCLUSION

Phonetics has been a key focus of linguistics research for many years, but its complexity has made it challenging to understand speech sounds accurately. To effectively address this, it is essential to stay updated with the latest technologies and incorporate them into the learning process. However, acquiring phonetic knowledge also presents difficulties for language learners due to the high demands of auditory discrimination and articulatory control. Therefore, further investigation and adaptation are needed for phonetics instruction to be more effective. The use of AI and AR technologies has shown promise in education, particularly in the teaching and learning of phonetics. These technologies offer personalized and interactive methods for instruction, allowing learners to receive adaptive feedback and correct their pronunciation in real time. This not only improves the learning experience but also promotes self-directed learning and retention of phonetic knowledge. By incorporating AI and AR, phonetics instruction can bridge the gap between theory and practice, making it more practical and relevant for language learning. This paper presented a new interactive AR system for learning phonetics. We demonstrated this system by creating a MAR app for the Arabic Phonetics Atlas textbook. The system provides a 3D animated model of speech organs and real-time feedback for students' pronunciation using a customized AI phoneme recognition system. We performed an extensive user study to evaluate the effectiveness and learning impact of the proposed system. For effectiveness evaluation, the quantitative data, in the form of ease of use and time to master the phoneme sound, were recorded for two groups. Each group has 40 participants. They learned the articulation of the/s/ sound in Arabic. One uses the traditional method, and the other uses our prototype. In addition, a comprehensive survey was conducted to evaluate the application's usability, impact on learning, and interactivity support. The survey results show that the system has a very high level of usability and supports interactivity. Students who used AR Atlas had a higher level of interaction and, consequently, learnt the phoneme sound faster. Finally, the learning impact survey demonstrated the effectiveness of the application as an interactive system. It justified its usage as a complementary tool with face-to-face traditional learning. Results showed that most participants struggled with pronouncing /s/ accompanied by a heavy sound, highlighting the importance of the ''Practice Words'' step in our system. The system outperformed traditional learning methods due to its interactive nature, real-time feedback, and animated illustrations. The proposed system can be utilized in other languages with intricate phonetic sounds, such as German and Chinese. Additionally, it may assist in effectively treating speech and language disorders, such as correcting lisp.

APPENDIX A
FIRST ARTICULATORY AWARENESS QUIZ

Questions
01. When you say/s/as in 'سما', do you raise your tongue to almost touch the alveolar ridge, the roof of the mouth?
02. When you say the word 'سورة', do your tongue protrude?
03. Does the following video clip show the correct tongue movement for the/s/ sound? 
04. Which of the audio files has /s/ sound?

APPENDIX B
SECOND ARTICULATORY AWARENESS QUIZ

Questions
01. When you say/s/as in 'سوط', do you raise your tongue high in the mouth to almost touch the alveolar ridge, the roof of the mouth?
02. Does your tongue protrude When you say the word 'سقف'?
03. Does the following video clip show the correct tongue movement for the/s/ sound? 
04. Which of the audio files has /s/ sound?

APPENDIX C
PRODUCTION TESTS

Pre-test			Post-test			Delayed test		
سهل	أسد	ناس	سهولة	اسلام	أناناس	سهم	أسرة	حارس
سلام	وسط	شمس	سجل	وساده	قوس	سجاد	وسام	طاووس
سماء	رسم	جرس	سحاب	فستان	درس	ساعه	أسود	ضرس

APPENDIX D
USABILITY AND IMPACT ON LEARNING QUESTIONNAIRE

No.	Questions
Q1	Previous experience with learning phonetics.
Q2	AR Atlas was able to teach the phonemes sound.
Q3	The steps to learn the/s/ sound using the AR Atlas app were clear.
Q4	I understood how to pronounce the/s/ sound.
Q5	It was easy to learn phonemes sound.
Q6	AR Atlas is more effective than the traditional Atlas textbook.
Q7	AR Atlas was more effective than the instructional video.
Q8	AR Atlas is interesting to use.
Q9	AR Atlas is easy to use.
Q10	I believe that using AR Atlas has contributed positively to my learning.
Q11	AR Atlas will completely replace the face-to-face learning method.
Q12	AR Atlas would be more suitable for the face-to-face learning method.
Q13	I'll keep practicing with the AR Atlas app.
Q14	I am satisfied with the live feedback.
Q15	I am satisfied with the graphics used.
Q16	AR Atlas app encouraged interaction.

REFERENCES

[1] C. Chang, "The phonetics of second language learning and bilingualism," *The Routledge Handbook of Phonetics Book*, Evanston, IL, USA: Routledge, 2019, ch. 15, pp. 427–447, doi: 10.4324/9780429056253-16.

[2] D. House, "Acoustic phonetics," in *The SAGE Encyclopaedia of Human Communication Sciences and Disorders*. AccessScience. Accessed: Jan. 11, 2024. [Online]. Available: https://doi.org/10.1036/1097-8542.802380

[3] B. Luo, "Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction," *Comput. Assist. Lang. Learn.*, vol. 29, no. 3, pp. 451–476, Apr. 2016, doi: 10.1080/09588221.2014.963123.

[4] *ELSA*. Accessed: Jan. 14, 2023. [Online]. Available: https://elsaspeak.com/en/product

[5] B. Zou, Y. Du, Z. Wang, J. Chen, and W. Zhang, "An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills," *SAGE Open*, vol. 13, no. 3, Aug. 2023, doi: 10.1177/21582440231193818.

[6] C. S. Che Dalim, M. S. Sunar, A. Dey, and M. Billinghurst, "Using augmented reality with speech input for non-native children's language learning," *Int. J. Hum.-Comput. Stud.*, vol. 134, pp. 44–64, Feb. 2020, doi: 10.1016/j.ijhcs.2019.10.002.

[7] A. Parmaxi and A. A. Demetriou, "Augmented reality in language learning: A state-of-the-art review of 2014–2019," *J. Comput. Assist. Learn.*, vol. 36, no. 6, pp. 861–875, Dec. 2020, doi: 10.1111/jcal.12486.

[8] A. Henrysson, "Bringing augmented reality to mobile phones," Dept. Sci. Technol., School Comput. Science Diss., Norrköping, no. 1145, 2007. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:16967/FULLTEXT01.pdf

[9] R. Tolba, T. Elarif, Z. Taha, and R. Hammady, "Mobile augmented reality for learning phonetics: A review (2012–2022)," in *Extended Reality and Metaverse, XR*. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-031-25390-4_7.

[10] G. Daniel, M. Moreira, and R. Mesquita, "InglesAR-a collaborative augmented reality game for English practicing," in *Proc. SBGames*, 2020, pp. 791–794.

[11] M. Fan, S. Sarker, and A. Antle, "From tangible to augmented: Designing a PhonoBlocks reading system using everyday technologies," in *Proc. CHI EA CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–6, Paper LBW555, doi: 10.1145/3170427.3188459.

[12] K. Thongchum and S. Charoenpit, "A conceptual design of kanji mobile application with augmented reality technology for beginner," in *Proc. 5th Int. Conf. Bus. Ind. Res. (ICBIR)*, May 2018, pp. 149–154.

[13] D. Sinyagovskaya and J. Murray, "Augmented reality in Chinese language pronunciation practice," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, 2021, pp. 403–408, doi: 10.1109/ISMAR-Adjunct54149.2021.00092.

[14] I. Nugraha, A. Suminar, D. Octaviana, M. Hidayat, and A. Ismail, "The application of augmented reality in learning English phonetics," *Phys. J., Conf. Ser.*, vol. 1402, May 2019, Art. no. 077024, doi: 10.1088/1742-6596/1402/7/077024.

[15] M. Aladin, A. Ismail, M. Salam, R. Kumoi, and A. Ali, "AR-to-KID: A speech-enabled augmented reality to engage preschool children in pronunciation learning," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 979, Nov. 2022, Art. no. 012011, doi: 10.1088/1757-899X/979/1/012011.

[16] A. R. Afriliani, Y. W. Efendi, and N. Pendit, "Developing augmented reality on English phonetics model," in *Proc. Int. Conf. Innov. Open Distance Learn. (INNODEL)*, vol. 3, 2022, pp. 121–129.

[17] W. A. A. W. Daud, M. T. A. Ghani, A. A. Rahman, M. A. B. M. Yusof, and A. Z. Amiruddin, "ARabic-kafa: Design and development of educational material for Arabic vocabulary with augmented reality technology," *J. Lang. Linguistic Stud.*, vol. 17, no. 4, pp. 1760–1772, Oct. 2021.

[18] *Mondly*. Accessed: Apr. 24, 2023. [Online]. Available: https://www.mondly.com/ar

[19] N. C. Hashim, N. A. A. Majid, H. Arshad, S. S. M. Nizam, and H. M. Putra, "Mobile augmented reality application for early Arabic language education-: Arabic," in *Proc. 8th Int. Conf. Inf. Technol. (ICIT)*, May 2017, pp. 761–766.

[20] *Quran-Tajweed (2005-2017)*. Accessed: Mar. 21, 2024. [Online]. Available: https://www.quran-tajweed.net/tagweed/index.php/19-pro-key/47-tajweed-dvd-2

[21] *Almaany (2010-2024)*. Accessed: Mar. 20, 2024. [Online]. Available: https://www.almaany.com/ar/dict/ar-ar/Al-Haruf-ul-Halqiyya/

[22] A. Chekayri, "Phonic approach in teaching reading in the 'Arabic language," Merrill-Palmer quarterly (Wayne), State Univ. Press, New York, NY, USA, Tech. Rep., 2018.

[23] N. Ahmed and M. Lataifeh, "Impact and analysis of a collaborative augmented reality educational environment," *J. Comput. Educ.*, Apr. 2023, doi: 10.1007/s40692-023-00275-x.

[24] R. M. Tolba, Z. T. Fayed, H. A. Alsayadi, N. O. Saad, and T. Elarif, "Detection and discrimination of Arabic phonemes using long short-term memory (LSTM) model," in *Proc. 11th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Nov. 2023, pp. 147–153, doi: 10.1109/ICICIS58388.2023.10391156.

[25] L. Davidson, "Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts," *J. Experim. Psychology: Human Perception Perform.*, vol. 37, no. 1, pp. 270–282, 2011, doi: 10.1037/a0020988.

[26] C. García and J. Mora, "Assessing the effects of phonetic training on L2 sound perception and production," in *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, 2009.

[27] S. Yan, D. Sun, Q. Zhou, Y. Yang, and P. Tian, "Exploring the impact of virtual laboratory with KWL reflective thinking approach on students' science learning in higher education," *J. Comput. Higher Educ.*, Sep. 2023, doi: 10.1007/s12528-023-09385-y.

[28] J. Zhu, X. Zhang, and J. Li, "Using AR filters in L2 pronunciation training: Practice, perfection, and willingness to share," *Comput. Assist. Lang. Learn.*, vol. 35, pp. 1–30, Jun. 2022, doi: 10.1080/09588221.2022.2080716.

[29] RLA College. *ANOVA (Analysis of Variance)*. Accessed: Feb. 14, 2024. [Online]. Available: https://rlacollege.edu.in/pdf/Statistics/Anova.pdf

[30] I. Saja, N. A. Mazlan, A. Fauzi, N. A. K. Zamri, N. A. M. Rohana, and A. A. Rapa, "Augmented reality in language learning: Enhancing student engagement in a third language subject," *Int. J. Academic Res. Bus. Social Sci.*, vol. 12, no. 8, pp. 1581–1593, Aug. 2022.

RAHMA M. TOLBA received the M.Sc. degree in computer science from Ain Shams University, Cairo, Egypt, in 2018, where she is currently pursuing the Ph.D. degree with the Department of Computer Science, Faculty of Computer and Information Sciences. Since 2019, she has been a University Lecturer. Her research interests include human–computer interaction, AI animation/graphics, extended reality, 3D animation, facial animation, and gamification.

TAHA ELARIF received the Doctor of Philosophy degree in computer science from the Université de Technologie de Compiègne (UTC), France, in 1983. Since 1979, he has been involved in teaching and conducting research. He is currently a Professor with the Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt. His research interests include computer graphics, image processing, artificial intelligence, and augmented reality.

ZAKI TAHA received the Ph.D. degree from the Faculty of Engineering, Ain Shams University, Cairo, Egypt, in 1996. He is currently a Professor with the Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University. Since 1984, he has been actively engaged in the field of computer science, with a focus on his research on speech processing and software engineering.

RAMY HAMMADY (Member, IEEE) received the Ph.D. degree from Staffordshire University, in 2019. Since 2016, he has been one of the pioneers of MS HoloLens. In 2019, he started to work at Solent University as a Lecturer in computer games. In 2021, he became a Senior Lecturer and the Department Head of Research in Games. He is an Assistant Professor with the School of Computer Science and Electronic Engineering. He is also a Microsoft Most Valuable Professional (MVP) in mixed reality, in addition to being a Researcher/Expert in AR/VR/MR technologies and a Game Designer/Developer. He has industrial outputs in the immersive systems and games domains. He has extensive experience using Microsoft HoloLens and other AR glasses. He has several publications in serious games, gamification, augmented, virtual, and mixed reality technologies in cultural heritage.