Interactive Web-Based Text-To-Speech and Speech Recognition Media for Enhancing Arabic Listening Proficiency

[1]Luthfiyatuz Zuhriyah
luthfieya929@gmail.com
Arabic Education Department, Faculty of Tabiyah and Teacher Training, Universitas Sains Al Qur'an, Indonesia

[2]Asep Sunarko *(Corresponding Author)*
asepsunarko@unsiq.ac.id
Arabic Education Department, Faculty of Tabiyah and Teacher Training, Universitas Sains Al Qur'an, Indonesia

[3]Ahmad Zuhdi
azuhdi@unsiq.ac.id
Islamic Religious Education Department, Faculty of Tabiyah and Teacher Training, Universitas Sains Al Qur'an, Indonesia

[4]Moh Ali Khusain
alikhusain63@gmail.com
Islamic Studies, Arrohuni Faqih Institut, Kenitra, Morocco

ABSTRACT

Background*:* Arabic listening lessons often rely on passive audio repetition, leaving students with little opportunity to check their understanding or receive immediate feedback. This condition slows their progress, especially in recognizing key phonemes and responding accurately.

Aims: This study aims to develop and validate a web-based interactive listening tool using text-to-speech (TTS) and speech-recognition (ASR) to support students' Arabic listening proficiency.

Methods: The research followed the Dick, Carey & Carey model through the stages of analysis, design, development, and expert validation. Data were collected through needs analysis, classroom observations, and interviews with teachers to identify learners' challenges in Arabic listening. After the prototype was created, expert judgment sheets were used to gather validation data from design and content specialists. The responses were then analyzed using percentage-based validity scoring to determine the feasibility of the media. The prototype integrated web-based TTS and ASR modules, specifying platform architecture, speech engines, supported languages, and technical parameters to ensure replicability.

Results: The results show that the media supports active listening practice and helps address common learner difficulties, particularly in phoneme recognition and comprehension. TTS provided clear and natural audio, while ASR offered direct corrective feedback. Expert assessments indicated high feasibility, with a design validity score of 92% and content validity of 87%.

Implications: The findings indicate that TTS–ASR integration can ease teachers' correction workload and help students practice more independently, offering practical potential for broader classroom use.

Keywords: Arabic listening; interactive web-based; text-to-speech; speech recognition media

1. INTRODUCTION

The rapid advancement of digital technology has significantly transformed the landscape of foreign language learning (Zhang & Dong, 2024), particularly in the domain of computer-assisted language learning (CALL) (Mahajan, 2022). In the context of Arabic as a foreign language (AFL), the integration of intelligent learning technologies remains relatively underdeveloped compared to other widely taught languages, despite the growing global demand for Arabic proficiency for academic, professional, and religious purposes (Al-Issa, 2020). Among the four core language skills, listening proficiency (Zhang & Graham, 2020), known as Mahārah Istimāʿ, plays a foundational role because it supports the development of speaking, reading, and writing (Jones, 2007). However, listening remains one of the most challenging skills for learners to acquire (Ghanipour & Bozorgian, 2025), especially in environments where exposure to native Arabic speakers is limited (Alandejani & Sayed, 2024). Traditional pedagogical approaches, which typically depend on audio recordings, teacher-led explanations, and limited in-class listening exercises, often fall short in providing sufficient repetition, personalised practice, and real-time feedback (Lee & Ko, 2025). Consequently, learners frequently struggle to decode authentic Arabic sounds (Zikrillah et al., 2025), identify essential vocabulary (Rakhlin et al., 2025), and interpret the meaning within connected speech (Itriq & Mohd Noor, 2025). These persistent difficulties highlight a critical need for innovative instructional solutions that utilise emerging technologies to enhance listening comprehension in a more engaging, interactive, and effective manner.

Against this background, the present research addresses a central issue arising from the gap between the growing need for individualised and intensive listening practice (Erdiana et al., 2025) and the limited capacity of traditional instructional methods to meet these demands (Bozorgian & Shamsi, 2025). Many learners of Arabic face challenges in distinguishing between similar phonemes, recognising prosodic patterns, and understanding contextual cues in oral texts (Asadi et al., 2025). These challenges are further compounded by the linguistic complexity of Arabic (Fanoush et al., 2025), which is characterised by rich morphology, unique phonetic features such as emphatic consonants, and significant distinctions between Modern Standard Arabic (MSA) and regional dialects. Moreover, most existing digital listening tools offer only linear, passive learning experiences, lacking the interactive features needed to promote engagement and self-regulated learning (Ardasheva et al., 2017). They rarely offer adaptive difficulty adjustments or immediate feedback beyond simple correctness indicators. As a result, listening activities remain largely instructor-dependent, leading to low motivation and insufficient improvement in learners' listening skills (Ebrahimzadeh & Ebadi, 2025). From this condition, the problem can be articulated as the insufficiency of current Arabic listening instruction tools in providing interactive, adaptive, and real-time feedback systems necessary to support meaningful enhancement of listening proficiency.

Recent developments in text-to-speech (TTS) and speech recognition (SR) technologies have opened new possibilities for language-learning innovations(Oumaima et al., 2018), particularly in the area of listening instruction (Dhonburi, et al., 2023). While these technologies have been widely integrated into learning platforms for English

and other major world languages (Kumar et al., 2023), Their application in Arabic learning environments remains limited. Previous research has demonstrated that TTS systems can generate high-quality, consistent Modern Standard Arabic audio, enabling learners to practice listening repeatedly, adjust the speed of delivery, and enhance their phonological awareness. Similarly, speech recognition technologies have shown promising potential in training learners to improve pronunciation accuracy and develop awareness of spoken linguistic forms (Morzy, 2025). However, most studies examine these technologies in isolation rather than integrating them into a holistic, interactive web-based environment. Additionally, earlier research rarely aligns the application of TTS and SR with pedagogical frameworks that support listening comprehension, such as bottom-up decoding, top-down processing, and integrative task design (Yan et al., 2025). Consequently, despite technological advancements, a clear gap persists in the development of integrated, interactive web-based TTS–SR media specifically designed to enhance Arabic listening proficiency.

This research responds to the pressing need for interactive, adaptive, and pedagogically sound digital tools that support the development of Arabic listening proficiency. By integrating text-to-speech and speech recognition technologies within a web-based environment, the study offers a comprehensive approach that fills existing gaps in the literature and provides practical solutions for learners and educators. The findings are expected to inform future developments in intelligent language learning systems, contributing to the ongoing advancement of CALL and enhancing Arabic language education across diverse learning contexts.

1.1 Research Gap and Novelty

A review of previous studies reveals several substantive gaps that this research seeks to address. Studies examining the use of text-to-speech (TTS) in Arabic listening instruction predominantly frame effectiveness in terms of technical output, such as voice naturalness, pronunciation clarity, and system performance, with limited attention to pedagogical integration. For instance, Fadillah and Bariyyah (Fadillah & Bariyyah, 2024). Explicitly position TTS as a tool for audio delivery enhancement, without embedding it within interactive listening tasks or providing mechanisms for learner feedback. Similarly, research on speech recognition (SR) in Arabic language learning has largely been confined to pronunciation and speaking practice, as reflected in prior works that evaluate SR accuracy at the segmental level but do not explore its application for assessing listening comprehension or supporting integrated listening activities (Mohammed, 2022). Although some recent platforms incorporate either TTS or SR, these systems remain functionally isolated and pedagogically limited. Chemnad and Othman (2023), for example, acknowledge that their proposed system focuses on individual components rather than a holistic listening pedagogy framework. Consequently, existing studies can be characterised as fragmented in scope, offering partial solutions without addressing listening as an interactive, feedback-driven process. To date, there is a notable absence of a comprehensive, web-based platform that systematically integrates both TTS and SR within an established listening pedagogy framework. This gap constitutes the central novelty of the present study and underscores the need for an instructional design that moves beyond technical feasibility toward pedagogically grounded, interactive listening support (Chemnad & Othman, 2023).

Given these considerations, the present research emerges as a timely response to the discrepancy between the availability of technological tools and the pedagogical needs for effective Arabic listening instruction. As educational institutions increasingly adopt digital and hybrid learning models, there is a growing demand for evidence-based

technological solutions that foster autonomous learning without compromising instructional quality (Aldhafiri, 2020; Mohd & Mikail, 2017; Ubaidillah et al., 2024). The integration of TTS and SR into a web-based platform provides opportunities to transform passive listening activities into dynamic, personalised learning experiences equipped with real-time feedback and learner performance tracking. Such a system not only aligns with the global movement toward AI-driven education but also addresses the specific challenges faced by Arabic learners, thereby fulfilling both pedagogical and technological imperatives.

1.2 Research Questions

The central problem addressed in this study concerns the inadequacy of current Arabic listening instruction, which still relies heavily on passive audio repetition, limited practice opportunities, and the absence of immediate corrective feedback. These limitations prevent students from effectively recognizing Arabic phonemes, understanding spoken texts, and responding accurately during listening activities. At the same time, teachers face constraints in providing individualized correction due to time and classroom management demands. Although text-to-speech (TTS) and speech recognition (SR) technologies have shown significant potential for supporting language learning in other major languages, their integration into Arabic listening pedagogy remains limited and underexplored.

Therefore, the main research question guiding this study focuses on the following points:

1. How can a web-based interactive learning media integrating text-to-speech (TTS) and speech recognition (SR) be designed to support the development of students' Arabic listening proficiency (*mahārah istimāʿ*)?
2. To what extent is the designed web-based TTS–SR interactive listening media valid in terms of content and instructional design, as evaluated by experts?
3. How can the validated TTS–SR-based interactive learning media be utilized to enhance students' Arabic listening proficiency (*mahārah istimāʿ*)?

2. METHODS
2.1 Research Design

This study adopts a Research and Development (R&D) approach using the instructional design model proposed by Dick, Carey, and Carey (Toker, 2022), as it provides a systematic and learner-centred framework that emphasizes needs analysis, instructional alignment, and formative evaluation. Unlike the 4D or Borg and Gall models, which focus more on product dissemination and large-scale field testing, the Dick, Carey, and Carey model is more suitable for developing and validating instructional media at the prototype level with a strong emphasis on instructional effectiveness. The model was selected because it offers a clear, systematic path for developing instructional media, from identifying learning needs to a comprehensive evaluation of the product's effectiveness. The focus of this research is the development of a web-based interactive learning tool that integrates Text-to-Speech (TTS) and Speech Recognition (SR) technologies to support students' Arabic listening skills (mahārah istimāʿ). By following the Dick and Carey framework, each stage of development from analyzing learners to producing and testing the media can be carried out in a structured and accountable manner. This design also allows the researcher to introduce improvements at several phases through formative evaluation, ensuring that the final product is both pedagogically sound and suitable for students in the MTs learning context in Wonosobo.

2.2 Research Subjects

The participants involved in this study were 45 eighth-grade students enrolled at an MTs in Wonosobo, Central Java. Their ages ranged between 13 and 14 years, and they had already been introduced to basic Arabic lessons in the previous academic period. Participants were selected using a purposive sampling technique, as the students were directly related to the instructional goals of the developed media. To ensure the product was properly tested, participants were divided into different stages of the formative evaluation. The one-to-one evaluation involved 3 students who represented varying levels of academic ability. A group of 10 students participated in the small-group trial, which aimed to examine usability and clarity of instructions. The largest group, comprising 32 students, participated in the field test, allowing the researcher to observe the implementation of the media in a more authentic classroom context. For the summative evaluation, all 45 students participated in evaluating the final version of the developed product.

2.3 Research Procedures

The development process was carried out through several stages based on the Dick, Carey, and Carey model. The first step was to identify the instructional goal, which was to improve students' Arabic listening comprehension in accordance with the MTs curriculum. The next step was an analysis of learners and the learning environment. This analysis included students' prior knowledge, their familiarity with digital tools, and common obstacles they face when listening to spoken Arabic. Based on this analysis, the researcher formulated specific performance objectives that students were expected to achieve, such as recognizing Arabic phonetic patterns, understanding short spoken texts, and responding appropriately to oral prompts.

Afterwards, the researcher designed assessment instruments aligned with the objectives. These instruments consisted of listening comprehension tests used as pre-tests and post-tests. Once the assessment tools were ready, a detailed instructional strategy was arranged. This strategy incorporated TTS to present accurate Arabic pronunciation and SR to allow students to interact with the media through spoken responses. The strategy also included guided listening tasks, repetition activities, and interactive quizzes designed to promote gradual improvement.

The next step was to develop web-based media using a modern web framework, integrating TTS and SR through external speech APIs to support Arabic audio delivery and response evaluation. The system included Arabic language models, and SR accuracy was tested and calibrated based on Indonesian learners' speech samples to ensure reliable feedback. The design emphasized clarity, accessibility, and automatic feedback mechanisms. Following the development phase, formative evaluations were conducted in three stages: one-to-one, small group, and field testing. Each stage provided different types of feedback, which were carefully analyzed and used to improve the media before producing the final version. The final step was the summative evaluation, conducted after revisions were completed to assess the product's overall effectiveness.

2.4 Research Instruments

Several instruments were used to collect data throughout the study. The primary instrument was a listening comprehension test, given before and after the intervention. This test contained multiple-choice and short-answer questions designed to measure

students' understanding of spoken Arabic. The test's content validity was reviewed by experts in Arabic language education. A questionnaire was also distributed to collect students' and teachers' responses regarding the quality, usability, and practicality of the web-based media. The instrument consisted of several statements for each evaluation component, and responses were measured using a Likert scale ranging from strongly disagree to strongly agree. To complement these instruments, an observation sheet was used during classroom implementation to record students' engagement, difficulties, and interaction with the system. Additionally, semi-structured interviews were conducted with selected participants and the Arabic teacher to obtain more detailed qualitative data related to students' experiences and suggestions.

2.5 Data Analysis

Both quantitative and qualitative approaches were used to analyse the data. The quantitative analysis began with normality testing to ensure that the pre-test and post-test scores met statistical assumptions, followed by a Paired Sample t-test to determine significant differences in students listening scores before and after using the web-based TTS–SR media, and an N-Gain calculation supported by descriptive statistics (mean, minimum, maximum, and standard deviation) to measure the magnitude of improvement. The feasibility of the developed media was further examined through expert validation, in which material and media experts rated the product using a Likert scale (1–5), and the scores were converted into percentages to determine the validity category. Meanwhile, qualitative data from questionnaires, interviews, and classroom observations were analyzed through thematic analysis involving data reduction, coding, theme categorization, and interpretation, with instrument validity and reliability strengthened through expert judgment and triangulation across multiple data sources.

3. FINDINGS AND DISCUSSION
3.1 Findings

In this section, the findings are presented and discussed in relation to the research questions that directed the course of this study; (1) How can a web-based interactive learning media integrating text-to-speech (TTS) and speech recognition (SR) be designed to support the development of students' Arabic listening proficiency (*mahārah istimāʿ*)? (2) To what extent is the designed web-based TTS–SR interactive listening media valid in terms of content and instructional design, as evaluated by experts? (3) How can the validated TTS–SR-based interactive learning media be utilized to enhance students' Arabic listening proficiency (*mahārah istimāʿ*)? The following discussion elaborates on each of these questions based on the data obtained from the design documentation, expert validation, and implementation findings.

Needs Analysis Findings

The analysis stage was conducted to identify the needs of students and teachers in listening learning at an MTs school in Wonosobo, Central Java. Based on the initial observation results, the listening learning process (*maharah istima'*) in the classroom was still passive, because teachers tended to rely on simple recordings or direct text reading without providing adequate opportunities for independent practice. The questionnaire results showed that most students had a high interest in using interactive digital media, while also experiencing difficulties in recognizing similar phonemes such as ح, هـ, ع, and أ. Teachers also expressed the need for media that could provide automatic feedback to overcome time constraints in correcting students' pronunciation individually. These findings indicate a clear need for technology-based learning media

that can provide consistent pronunciation models and direct evaluation of oral responses.

Media Design Result

The design stage focused on developing the learning flow, formulating objectives, and designing the media interface. This media is designed to help students improve their ability to recognize phonemes, understand short utterances, and produce verbal responses as instructed. The learning structure is organized into several modules, ranging from phoneme recognition using TTS, word and sentence repetition exercises with SR feedback, to comprehension exercises. The interface is designed to be simple and responsive, making it easy to access on laptops and mobile devices. In addition, the technologies used include the Web Speech API for TTS and SR integration, and HTML, CSS, and JavaScript as the foundation for development. This initial design was then submitted to validators for feedback before proceeding to the development stage.
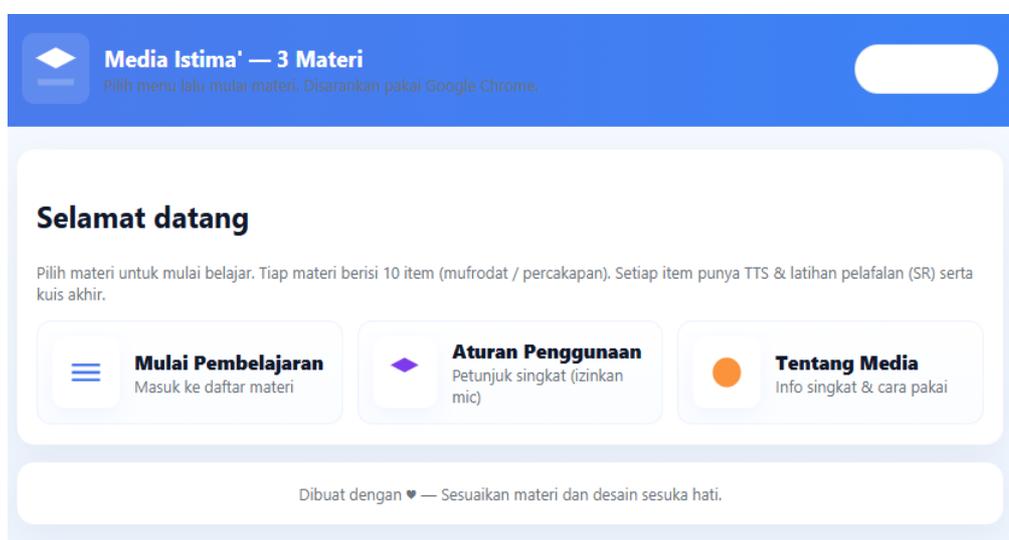
Figure 1. Homepage of the Web-Based *Istimā'* Learning Media

Product Development Result

During development, the initial design was transformed into a functional, fully interactive web-based learning medium integrating Text-to-Speech (TTS) and Speech Recognition (SR). The development process involved finalizing the layout, adjusting the colour scheme for each material category, refining typography, and ensuring visual consistency across all pages. The design emphasized clarity, simplicity, and ease of navigation, which are essential for MTs learners. The first component developed was the homepage, which serves as the main navigation hub. This page displays the welcome message and three primary menu options: Learning Menu, Usage Instructions, and About the Media. The layout is responsive and works on both mobile and desktop browsers.

Figure 2. Material Selection Page Featuring Three Topics

Next, the Material Selection Page was developed to allow learners to choose among the three available topics: (1) التَّعَارُفْ – *Perkenalan* (Introduction), (2) هِوَايَةٌ – *Hobi* (Hobies), and (3) حِوَارٌ - *Percakapan Pendek* (Short Conversation). Each topic contains 10 learning items consisting of mufrodat or short dialogues. This page also uses colour-coded icons to help students easily distinguish between materials.

The final and core component developed was the Learning Page, which integrates Text-to-Speech (TTS) and Speech Recognition (SR) features into a single interactive slide. Each learning item displays Arabic text along with an optional Indonesian translation. Learners can listen to the pronunciation through the TTS button and subsequently practice saying the phrase using the SR button.

When students press the Record (SR) button, the system analyses their pronunciation and generates real-time recognition output. If the student's pronunciation matches the target utterance, the system automatically awards 1 point, which is accumulated in the Material Score indicator displayed on the right side of the interface. This scoring mechanism provides immediate feedback and motivates learners to continue practising pronunciation until they achieve accurate output. The integration of TTS, SR, and the automatic scoring system makes the listening and pronunciation activities more interactive, adaptive, and supportive of autonomous learning.
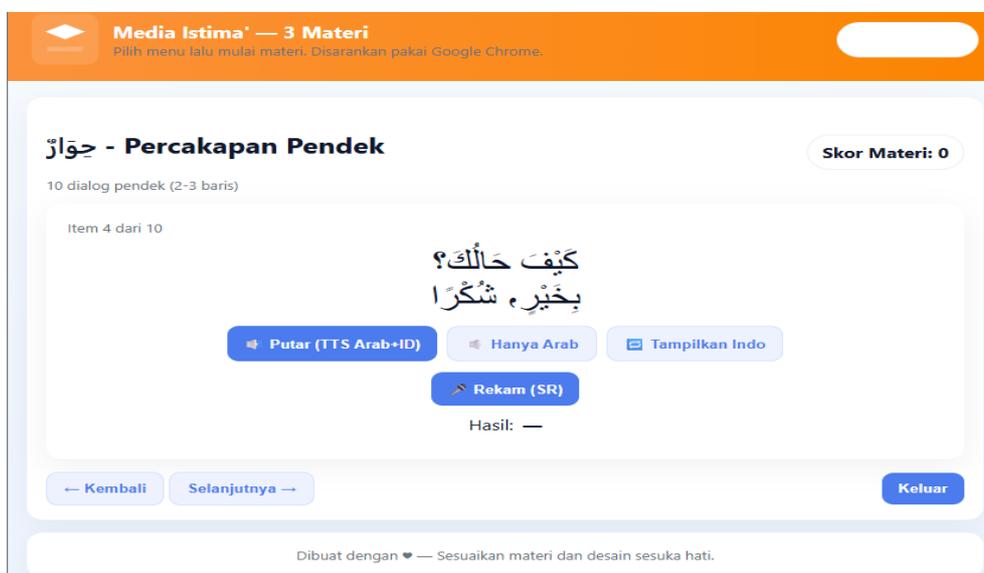
Figure 3. TTS–SR Learning Page: Displaying a Short Dialogue

Expert Validation

The learning media developed was validated by two experts: a learning design expert and an Arabic language materials expert. The instructional designer assessed the interface, navigation structure, relevance of features, clarity of flow, and suitability of the media to the characteristics of students at the MTs level. Meanwhile, the Arabic language expert assessed the accuracy of the content, the suitability of the material to the basic competencies, the quality of pronunciation produced by the TTS, and the relevance of the exercises to the learning objectives of *mahārah istimā'*.

The validation instrument used a 1–5 Likert scale, and the scores were converted to percentages to determine the product's feasibility. Based on the assessment results, the design expert gave a score of 92% (very valid category), indicating that the interface layout, navigation flow, and learning structure were in accordance with instructional design principles. Meanwhile, the content expert gave a score of 87% (highly valid category), indicating that the Arabic content, the quality of the TTS audio presentation, and the suitability of the exercises for the basic *mahārah istimā'* competencies met the required standards. Overall, the validation results indicate that this web-based learning media is suitable for use, with minor revisions based on expert input.

Table 1 Expert Validation Summary

Validator	Score (%)	Category
Design Expert	92%	Very valid
Content Expert	87%	Very valid

Quantitative Findings

The results of the pre-test and post-test on students' listening skills from one MTs in Wonosobo, Central Java, are presented in Table 2. The average pre-test score shows that students' initial skills remained low, while the average post-test score was higher after using TTS-SR-based learning media. The differences in the minimum, maximum, and standard deviation scores indicate a change in the distribution of scores before and after the treatment, as discussed in the discussion section.

Table 2 Descriptive Statistics of Pre-test and Post-test

Test Type	N	Std. Deviation	Minimum	Maximum
Pre-Test	45	8.71	41	75
Post-Test	45	7.95	60	92

The results show an increase of 19.71 points from the initial test to the final test, reflecting an overall improvement in students' listening skills after using the learning media.

Normality Test

The Shapiro–Wilk test was used to determine whether the test scores were normally distributed.

Table 3 Shapiro–Wilk Normality Test

Test	Statistic	df	p-value
Pre-Test	0.965	45	0.088
Post-Test	0.972	45	0.113

Both tests produced p-values greater than 0.05, indicating that the data were normally distributed.

Homogeneity Test
Levene's test was conducted to verify homogeneity of variance.

Table 4 Levene's Homogeneity Test

Test	F	df1	df2	p-value
Pre-Test vs Post-Test	1.86	1	88	0.176

Since $p > 0.05$, the data fulfilled the homogeneity requirement.

Paired Sample T-test
A paired-samples t-test was conducted to determine the difference between pre-test and post-test scores. The results shown in Table 5 indicate a p-value of < 0.05, indicating a significant difference in scores after the use of web-based learning media.

Table 5 Paired Sample T-Test Result

Comparison	t-value	df	Sig. (p)	Conclusion
Pre vs Post	11.27	44	0.000	Significant

Effectiveness Analysis (N-Gain)
To measure the effectiveness of the intervention, the N-Gain score was calculated.

Table 6 N-Gain Effectiveness Index

Pre-Test Mean	Post-Test Mean	N-Gain	Category
58.42	78.13	0.47	Moderate

Based on the N-Gain of 0.47, the learning media demonstrated moderate effectiveness in improving listening proficiency.

Qualitative Findings
Web-based media with text-to-speech (TTS) and speech recognition (SR) gives students a new way to learn Arabic listening skills. From the start, students showed strong enthusiasm, seeing this media as a distinct change from their previous experiences. Many said that digital voices, like those of native speakers, helped them better understand Arabic pronunciation. They could grasp intonation, rhythm, and vocal stress more easily compared to teacher explanations or old lesson recordings.

When using the TTS feature, students replay audio as often as needed, deepening their understanding of Arabic sounds at their own pace. Students aware of weaknesses in distinguishing phonemes with similar articulation found TTS helpful for its clear,

consistent voice. They reported improved recognition of sound differences after several practice sessions. This media supports independent learning because the materials are always accessible, allowing students to control their learning pace.

The speech recognition feature gives students a unique experience. They feel as if they have a virtual teacher who evaluates their pronunciation in real time. When a student says a word or a sentence, the system provides immediate feedback in the form of a score or an accuracy indicator. Most students say this motivates them to repeat exercises until they achieve good results. This feedback makes them more aware of articulation errors and helps them correct mistakes on their own. Students feel more comfortable practicing with SR than in front of the class because they avoid embarrassment or fear of peer evaluation. Besides improving listening skills, this web-based media boosts motivation and engagement. The environment becomes more interactive and less boring. Digital features make Arabic learning feel modern and enjoyable, not just limited to lectures or repetitive listening. Students tend to follow instructions and practice more on their own, even outside class. Some say the media challenge them, turning practice into a game that tests pronunciation.

Student responses showed several obstacles to using these media. Poor internet access was a major obstacle, especially in areas with unstable networks. Connection disruptions often prevented the speech recognition feature from working, leading to inaccurate evaluations. Students felt confused or frustrated when the system did not capture their voices accurately. Some students also needed more time to adapt to this technology, especially those unfamiliar with web-based applications. At first, they felt awkward and needed guidance to use the features.

Despite these obstacles, most students responded positively to TTS and SR media. They valued easy access, flexible learning times, and practicing without pressure. Many said these media increased their confidence in understanding and pronouncing Arabic sounds. The tools significantly helped students internalize Arabic sounds, especially those who previously felt they had little exposure. Teachers observed students becoming more independent and active in evaluating their development, which changed classroom dynamics. Overall, students saw web-based media with TTS and SR as a significant improvement. The technology-enhanced learning opened up new ways to understand listening material and increased motivation, independence, and engagement in Arabic learning.

3.2 Discussion

This section discusses the key findings in relation to established theories and prior research on technology-enhanced Arabic listening instruction. It focuses on how the integration of Text-to-Speech (TTS) and Speech Recognition (SR) addresses learners' listening difficulties and adds pedagogical value in Arabic as a Foreign Language (AFL) contexts. Three main aspects are discussed: the role of consistent TTS input in phonological decoding, the contribution of SR feedback to learner engagement and self-regulation, and the added value of integrating TTS and SR within a task-oriented listening framework.

Enhancing Phonological Decoding through Consistent TTS Input

The findings indicate that consistent exposure to TTS input improved students' ability to recognise similar Arabic phonemes. The stable pronunciation provided by the TTS engine enabled repeated listening without variability, supporting bottom-up processing by strengthening sound–meaning mapping. This aligns with (Sun, 2023) and (Alsuhaibani et al., 2024), who reported that repeated digital pronunciation input enhances phonological awareness. The controlled listening environment also appeared

to reduce cognitive load and improve learner focus.

Fostering Engagement and Self-Regulation via Interactive SR Feedback

The SR component fostered learner engagement by providing immediate, non-judgmental feedback. Initial reluctance to speak decreased as learners repeatedly interacted with the system, supporting self-regulated learning through monitoring and self-correction. This finding is consistent with (Rehman et al., 2024) and (Wang et al., 2024), who showed that automatic feedback promotes learner autonomy and motivation. Learners who engaged more frequently with SR activities demonstrated faster progress.

The Synergistic Value of Integrated TTS–SR in AFL Listening Pedagogy

The integration of TTS and SR offered added pedagogical value by transforming passive listening into an interactive cycle of input, output, and feedback. This approach aligns with task-based language teaching (TBLT) principles, emphasising meaningful interaction and learner engagement. Although some learners initially faced challenges due to limited digital literacy, as noted by (Chalghoumi et al., 2022), (Hill & Hannafin, 2001), and (Chen, 2020) The overall findings suggest that integrated TTS–SR environments provide flexible, learner-centred listening practice that surpasses traditional teacher-centred approaches.

4. CONCLUSION

This study set out to design and evaluate an interactive web-based listening tool that integrates text-to-speech and speech-recognition technologies to support students' Arabic listening proficiency. Across the stages of analysis, design, and development guided consistently by the Dick, Carey & Carey model, the research showed that many of the barriers' students face in listening activities stem from limited exposure, a lack of immediate feedback, and the passive learning formats typically used in class. The media developed in this project offers a more active alternative, allowing learners to listen, respond, and receive instant feedback in a cycle that encourages gradual improvement. The expert validations, both in design and content, confirmed the platform's feasibility and pedagogical soundness, marking its contribution not only as a practical solution for classrooms but also as an example of how TTS and ASR can be meaningfully integrated into foreign-language listening instruction.

Even so, the research is not without limitations. The speech-recognition feature still needs refinement, especially for phonemes commonly mispronounced by beginners, and the system's reliance on a stable internet connection limits its use in certain environments. Future work may explore improving the recognition engine, incorporating a wider range of listening texts, or developing an offline or hybrid version of the media. From a practical standpoint, this study highlights that technology-supported listening tools can ease teachers' feedback workload and enable students to practice more independently. With continued development, the platform has strong potential to enhance the overall quality of Arabic listening instruction and inspire further innovations in digital language learning.

also like to thank the MTs in Wonosobo, Central Java, for granting permission and providing opportunities for researchers to conduct product trials in mahārah istimāʿ learning. We would like to express our deepest appreciation to the learning design validators and Arabic language experts for their valuable input, evaluation, and constructive suggestions in refining the web-based media developed. Appreciation is also extended to all students who actively participated in the one-on-one, small group, and field-testing stages, enabling this research to be carried out successfully and produce meaningful findings for the development of technology-based Arabic language learning.

Authors' Contributions

Luthfiyatuz Zuhriyah contributed to initiating and designing the study, conducting needs analysis, developing the conceptual design of the learning media, leading the development of the web-based application integrating Text-to-Speech (TTS) and Speech Recognition (SR) technologies, designing the interface, learning flow, and interactive features, collecting data, performing quantitative and qualitative data analysis, interpreting research results, and compiling the Introduction, Findings, Discussion, and Conclusion sections. Asep Sunarko contributed to compiling the background of the problem, formulating the research problem conceptually and systematically, strengthening the theoretical framework and alignment between research problem, objectives, and scientific contributions, supervising the research process, and critically reviewing the manuscript to ensure coherence and academic quality. Ahmad Zuhdi contributed to developing and validating research methods, including designing the Dick, Carey & Carey model-based Research and Development (R&D), determining research procedures, developing research instruments, technically validating experimental designs and data analysis, and refining the Research Methods section to ensure compliance with educational research standards. Moh Ali Khusain contributed to creating, refining, and testing the audio quality of the Text-to-Speech feature, providing input on articulation clarity, phonetic quality, and pronunciation consistency according to Arabic language standards, performing technical reviews of audio aspects, and providing final evaluations of product readiness before field implementation.

AI Generative Statement

In preparing this manuscript, the authors used generative artificial intelligence only in a limited capacity, primarily to assist with grammar refinement and language polishing. The conceptual framework, research methodology, data analysis, and interpretation of findings remain fully the intellectual work and responsibility of the authors. The authors affirm that the incorporation of AI tools did not compromise the integrity, originality, or scientific validity of the study.

REFERENCES

Alandejani, J. A., & Sayed, G. (2024). The implementation of high-impact practices and communication competency using the Arabic language. *Cogent Education, 11*(1), 2401253. https://doi.org/10.1080/2331186X.2024.2401253

Aldhafiri, M. D. (2020). The effectiveness of using interactive white boards in improving the Arabic listening skills of undergraduates majoring in Arabic language at Kuwaiti universities. *Education and Information Technologies, 25*(5), 3577–3591. https://doi.org/10.1007/s10639-020-10107-5

Al-Issa, A. S. M. (2020). The language planning situation in the Sultanate of Oman. *Current Issues in Language Planning, 21*(4), 347–414.

https://doi.org/10.1080/14664208.2020.1764729

Alsuhaibani, Y., Mahdi, H. S., Al Khateeb, A., Al Fadda, H. A., & Alkadi, H. (2024). Web-based pronunciation training and learning consonant clusters among EFL learners. *Acta Psychologica, 249*, 104459. https://doi.org/10.1016/j.actpsy.2024.104459

Ardasheva, Y., Wang, Z., Adesope, O. O., & Valentine, J. C. (2017). Exploring effectiveness and moderators of language learning strategy instruction on second language and self-regulated learning outcomes. *Review of Educational Research, 87*(3), 544–582. https://doi.org/10.3102/0034654316689135

Asadi, I. A., Kawar, K., & Tarabeh, G. (2025). Development of verb and noun word patterns in Arabic: A comparison between typically developing children and those with reading difficulties. *Journal of Speech, Language, and Hearing Research, 68*(8), 3976–3988. https://doi.org/10.1044/2025_JSLHR-24-00673

Bozorgian, H., & Shamsi, E. (2025). A review of research on metacognitive instruction for listening development. *International Journal of Listening, 39*(1), 1–16. https://doi.org/10.1080/10904018.2023.2197008

Chalghoumi, H., Al-Thani, D., Hassan, A., Hammad, S., & Othman, A. (2022). Research on older persons' access and use of technology in the Arab region: Critical overview and future directions. *Applied Sciences, 12*(14), 7258. https://doi.org/10.3390/app12147258

Chemnad, K., & Othman, A. (2023). Advancements in Arabic text-to-speech systems: A 22-year literature review. *IEEE Access, 11*, 30929–30954. https://doi.org/10.1109/ACCESS.2023.3260844

Chen, C.-H. (2020). Impacts of augmented reality and a digital game on students' science learning with reflection prompts in multimedia learning. *Educational Technology Research and Development, 68*(6), 3057–3076. https://doi.org/10.1007/s11423-020-09834-w

Dhonburi Rajabhat University, Thailand, Prasongngern, P., Soontornwipast, K., & Ed.D., Language Institute, Thammasart University, Thailand. (2023). Effects of listening strategy instruction incorporating intensive and extensive listening on listening skills and metacognitive awareness. *International Journal of Instruction, 16*(4), 155–172. https://doi.org/10.29333/iji.2023.16410a

Ebrahimzadeh, Y., & Ebadi, S. (2025). An exploration into the development of Iranian EFL learners' oral fluency through online dynamic assessment: A case study. *Cogent Education, 12*(1), 2549788. https://doi.org/10.1080/2331186X.2025.2549788

Erdiana, L., Dziqy, A. N. A., Farouq, A. A., & Slamet, J. (2025). Enhancing listening comprehension in non-English majors through AI-integrated gamified formative assessment. *Applied Research on English Language, 14*(3). https://doi.org/10.22108/are.2025.144695.2475

Fadillah, R., & Bariyyah, N. A. (2024). Implementation of the use of artificial intelligence (AI) based animation media to enhance vocabulary learning skills of elementary school students. *SHS Web of Conferences, 202*, 06004. https://doi.org/10.1051/shsconf/202420206004

Fanoush, T., Al-Khatib, W. G., Amro, M., Alzahrani, A., & Elshafei, M. (2025). Mispronunciation detection and diagnosis for young Arabic learners using transfer learning. *IEEE Access, 13*, 175047–175068. https://doi.org/10.1109/ACCESS.2025.3616335

Ghanipour, F., & Bozorgian, H. (2025). EFL learners' challenges in metacognitive listening intervention. *International Journal of Applied Linguistics, ijal.12862*. https://doi.org/10.1111/ijal.12862

Hill, J. R., & Hannafin, M. J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37–52. https://doi.org/10.1007/BF02504914

Itriq, M., & Mohd Noor, M. H. (2025). Arabic hate speech detection using deep learning: A state-of-the-art survey of advances, challenges, and future directions (2020–2024). *PeerJ Computer Science, 11*, e3133. https://doi.org/10.7717/peerj-cs.3133

Jones, D. (2007). Speaking, listening, planning and assessing: The teacher's role in developing metacognitive awareness. *Early Child Development and Care, 177*(6–7), 569–579. https://doi.org/10.1080/03004430701378977

Kumar, Y., Koul, A., & Singh, C. (2023). Deep learning approaches in text-to-speech system: A systematic review and recent research perspective. *Multimedia Tools and Applications, 82*(10), 15171–15197. https://doi.org/10.1007/s11042-022-13943-4

Lee, J., & Ko, Y. (2025). Comparing metaverse and face-to-face instruction for enhancing English skills and attitudes among EFL adolescents. *Education and Information Technologies, 30*(6), 8215–8244. https://doi.org/10.1007/s10639-024-13162-4

Mahajan, M. (2022). A literature review of the opportunities and barriers in computer-assisted language learning (CALL). *Journal of English Language and Literature, 09*(03), 37–48. https://doi.org/10.54513/joell.2022.9305

Mohammed, T. (2022). Designing an Arabic speaking and listening skills e-course: Resources, activities, and students' perceptions. *Electronic Journal of E-Learning, 20*(1), 53–68. https://doi.org/10.34190/ejel.20.1.2177

Mohd Sidek, H., & Mikail, I. (2017). Arabic as a second language listening comprehension: Instruction and assessment. *Ulum Islamiyyah, 20*, 21–34. https://doi.org/10.33102/uij.vol20no0.32

Morzy, M. (2025). *Spoken language processing: Conversational AI for spontaneous human dialogues* (1st ed.). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-88566-2

Oumaima, Z., Abdelouafi, M., & Meryem, E. H. (2018). Text-to-speech technology for Arabic language learners. In *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)* (pp. 432–436). https://doi.org/10.1109/CIST.2018.8596372

Rakhlin, N. V., Li, N., Aljughaiman, A., & Grigorenko, E. L. (2025). "Speech is golden": The importance of colloquial Arabic for reading standard Arabic for beginning readers. *Journal of Speech, Language, and Hearing Research, 68*(3S), 1441–1467. https://doi.org/10.1044/2024_JSLHR-23-00522

Rehman, N., Huang, X., Mahmood, A., Maqbool, S., & Javed, S. (2024). Enhancing the quality of research synopsis of international students through peer feedback: A case study. *New Directions for Child and Adolescent Development, 2024*(1), 1271802. https://doi.org/10.1155/2024/1271802

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology, 14*, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Toker, S. (2022). The progress of 21st-century skills throughout instructional design projects: A quasi-experimental comparison of rapid prototyping and Dick and Carey models. *Education and Information Technologies, 27*(2), 1959–1992. https://doi.org/10.1007/s10639-021-10673-2

Ubaidillah, U., Millah, F. I., & Sapitri, N. (2024). The use of online media "alefbata.com" in improving Arabic listening skills: Experimental study. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya, 12*(1), 103–114. https://doi.org/10.23971/altarib.v12i1.7852

Wang, X., Gao, Y., Sun, F., & Wang, Q. (2024). Unveiling the tapestry of teacher belief research: Tracing the present and forging the future through bibliometric analysis. *Current Psychology, 43*(17), 15659–15672. https://doi.org/10.1007/s12144-023-

05546-5

Yan, W., Li, B., & Lowell, V. L. (2025). Integrating artificial intelligence and extended reality in language education: A systematic literature review (2017–2024). *Education Sciences, 15*(8), 1066. https://doi.org/10.3390/educsci15081066

Zhang, P., & Graham, S. (2020). Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning, 70*(4), 1017–1053. https://doi.org/10.1111/lang.12411

Zhang, Y., & Dong, C. (2024). Exploring the digital transformation of generative AI-assisted foreign language education: A socio-technical systems perspective based on mixed-methods. *Systems, 12*(11), 462. https://doi.org/10.3390/systems12110462

Zikrillah, Erlina, E., Rafli, Z., & Amrulloh, M. A. (2025). The contribution of bilingualism to the enhancement of Arabic listening and speaking skills in language instruction. *Jurnal Pendidikan Bahasa, 14*(1), 183–200. https://doi.org/10.31571/bahasa.v14i1.9058