AN NABIGHOH

Journal homepage: http://e-journal.metrouniv.ac.id/an-nabighoh

ELEVENLABS TEXT-TO-SPEECH AND ARABIC LISTENING ACHIEVEMENT IN PESANTREN

Asep Sunarko [a,*], Muhamad Solehudin [b], Nurin Sakinah [c]

[a] Department of Arabic Language Education, Faculty of Education and Teacher Training, Universitas Sains Al Qur'an, Wonosobo, Indonesia
[b] Department of Arabic Language Education, Faculty of Education, Universitas Islam Internasional Darullughah Wadda'wah, Pasuruan, Indonesia
[c] Department of Arabic Language and Literature, Faculty of Islamic Revealed Knowledge and Human Sciences, International Islamic University Malaysia, Kuala Lumpur, Malaysia

Article Info

***Corresponding Author:**
Name:
Asep Sunarko
Email:
asepsunarko@unsiq.ac.id

Article History:
Received: October 22, 2025
Revised: March 3, 2026
Accepted: March 27, 2026
Published: April 6, 2026

Abstract

Background: Listening comprehension remains a major challenge in Arabic language learning, particularly in non-Arabic-speaking contexts such as Indonesian pesantren. Limited exposure to authentic auditory input often hinders phoneme recognition, prosodic awareness, and comprehension of connected speech. Recent advances in artificial intelligence, especially Text-to-Speech (TTS) technology, offer new opportunities to provide consistent, high-quality auditory input that can enhance listening skill development. **Research Objectives:** This study aims to examine the effectiveness of the ElevenLabs Text-to-Speech technology in improving students' Arabic listening comprehension and to explore students' perceptions regarding the clarity, authenticity, and motivational impact of AI-generated Arabic audio. **Methodology:** The study employed a mixed-methods quasi-experimental pre-test–post-test design complemented with interviews and classroom observations. Sixty intermediate students from Pondok Pesantren Darullughah Wadda'wah were assigned to an experimental group and a control group. The experimental group used ElevenLabs TTS materials, while the control group used teacher-narrated audio. Quantitative data were analyzed using paired-sample t-tests and qualitative data were processed through thematic analysis. **Results:** The findings revealed a significant improvement in the experimental group's listening scores, increasing from a mean of 63.2 to 82.7, while the control group showed only moderate improvement from 64.5 to 71.4. The statistical analysis indicated a significant difference ($p < 0.05$). Qualitative results also showed that students perceived the AI-generated voices as clearer, more authentic, and emotionally engaging, which enhanced their motivation and reduced listening anxiety. **Unique Contribution:** This study contributes to the emerging field of AI-assisted Arabic language education by demonstrating how human-like TTS technology can be effectively integrated into pesantren-based learning environments while maintaining both pedagogical and spiritual values. **Conclusion:** The study confirms that ElevenLabs TTS significantly enhances Arabic listening comprehension by providing consistent, expressive, and cognitively accessible auditory input for learners. **Recommendations:** Future studies are recommended to explore long-term applications of AI-based listening tools across other Arabic language skills and in broader Islamic educational contexts.

Keywords: ElevenLabs; Text-to-Speech; Arabic Listening Skills; Artificial Intelligence; Pesantren Education.

Please cite this article as Asep Sunarko, Muhamad Solehudin, and Nurin Sakinah.
"Elevenlabs Text-To-Speech and Arabic Listening Achievement in Pesantren."
An Nabighoh 28, no. 1 (2026): 1-20. DOI: https://doi.org/10.32332/an-nabighoh.v28i1.1-20.

Introduction

In the contemporary landscape of Arabic language education, listening remains one of the most challenging skills to master, particularly within non-Arabic-speaking contexts such as Indonesian pesantren.[1] Despite the strong emphasis on grammar (nahwu) and morphology (ṣarf), listening comprehension (maharah al-istimā') is often underdeveloped due to limited exposure to authentic, fluent, and emotionally natural Arabic input.[2] The reliance on teacher narration or prerecorded materials with artificial intonation frequently leads to low auditory engagement and shallow comprehension.[3] Consequently, many students struggle to differentiate phonemes, decode connected speech, or internalize natural prosody skills that are essential for real communicative competence in Arabic.[4]

In practical classroom settings at pesantren, these challenges manifest in observable learning difficulties. Preliminary classroom observations conducted prior to this study revealed that many students were unable to accurately transcribe short Arabic audio passages, frequently misidentified phonemic contrasts such as /ḥ/ and /h/ or /ṣ/ and /s/, and demonstrated hesitation when responding to comprehension questions delivered at natural speech rates. Teachers reported that listening sessions often resulted in passive note-taking rather than active meaning construction, with students requesting repeated teacher explanations instead of independently processing auditory input. Furthermore, the limited availability of varied and high-quality Arabic audio resources restricted students' exposure to diverse speech patterns, resulting in overreliance on a single teacher's vocal style. These field-based conditions indicate that the issue is not merely theoretical but pedagogically urgent, requiring an intervention that can provide consistent, natural, and repeatable auditory input within the pesantren learning environment.

In addition, exposure to authentic native Arabic speech remains limited in many pesantren, particularly in rural areas where interaction with native speakers is rare. Students are often accustomed to localized instructional accents, which may not fully reflect the natural prosodic variation of contemporary Arabic. ElevenLabs-generated audio provides access to high-fidelity, native-like pronunciation and prosodic consistency that are seldom available in under-resourced pesantren environments. Such exposure may broaden learners' phonological sensitivity and enhance their familiarity with standardized Arabic speech patterns, thereby reducing reliance on a single instructional voice model.

---

[1] Yusuf Arisandi and Moh. Tohiri Habib, "Optimizing YouTube for Interactive Arabic Learning in Pesantren: Effective Content Creation Strategies," *International Journal of Arabic Language Teaching* 7, no. 02 (2025): 239–54, https://doi.org/10.32332/ijalt.v7i02.10363.

[2] Dana Bsharat-Maalouf et al., "The Involvement of Listening Effort in Explaining Bilingual Listening Under Adverse Listening Conditions," *Trends in Hearing* 27 (January 2023): 23312165231205107, https://doi.org/10.1177/23312165231205107.

[3] Yogesh Kumar et al., "A Deep Learning Approaches in Text-to-Speech System: A Systematic Review and Recent Research Perspective," *Multimedia Tools and Applications* 82, no. 10 (2023): 15171–97, https://doi.org/10.1007/s11042-022-13943-4.

[4] Maysa Ahmad et al., "Assessing AI-Driven Dubbing Websites: Reactions of Arabic Native Speakers to AI-Dubbed English Videos in Arabic," *Research Journal in Advanced Humanities* 6, no. 1 (2025), https://royalliteglobal.com/advanced-humanities/article/view/1963.

From a psycholinguistic perspective, listening comprehension constitutes a dynamic interaction between perceptual decoding, working memory, attentional control, and affective regulation. Research in second language acquisition increasingly recognizes that learners' listening difficulties are not solely caused by linguistic complexity but also by unstable input conditions, inconsistent pacing, and heightened anxiety during real-time auditory processing. In this regard, AI-based Text-to-Speech systems offer a form of pedagogically sustainable input, allowing learners to repeatedly access stable, high-fidelity auditory material without temporal pressure. Such stability supports deeper phonological mapping, strengthens auditory memory traces, and facilitates gradual automatization of listening skills. Moreover, when AI-generated voices incorporate emotional contour and natural prosody, they activate learners' affective engagement, which has been shown to enhance attention and retention.[5] Consequently, the pedagogical value of advanced TTS technologies lies not only in their technical accuracy but also in their capacity to create a cognitively safe and emotionally responsive listening environment—an aspect that is particularly crucial in Arabic learning contexts where phonological density and rhythmic patterns pose persistent challenges for non-native learners.[6]

The advent of Artificial Intelligence (AI) and advanced Text-to-Speech (TTS) systems has transformed the way listening input can be delivered.[7] Among these innovations, ElevenLabs TTS has emerged as a leading technology capable of producing highly natural, human-like Arabic speech with authentic emotional tone and prosodic variation.[8] Unlike conventional TTS engines that often produce robotic or monotonic voices, ElevenLabs integrates deep-learning-based speech synthesis that mimics human rhythm, emotion, and intonation. This innovation holds immense potential for Arabic education in pesantren, where authentic input and emotional engagement are crucial for cultivating listening sensitivity and a sense of spiritual resonance during language learning.

[5] Shelley Xiuli Tong et al., "How Prosodic Sensitivity Contributes to Reading Comprehension: A Meta-Analysis," *Educational Psychology Review* 35, no. 3 (2023): 78, https://doi.org/10.1007/s10648-023-09792-8; Jun Chen and Xinran Lehto, "The Impact of Sound Design with AI Synthetic Voices on the Listening Experience in Audio Tour Guides," *Information Technology & Tourism* 27, no. 4 (2025): 1081–109, https://doi.org/10.1007/s40558-025-00332-4.

[6] Ziyang Ma et al., "Leveraging Speech PTM, Text LLM, And Emotional TTS For Speech Emotion Recognition," *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 14, 2024, 11146–50, https://doi.org/10.1109/ICASSP48485.2024.10445906.

[7] Dedi Mulyanto et al., "Utilization of Artificial Intelligence with Text-To-Speech Technology Based on Natural Language Processing to Enhance Arabic Listening Skills for Non-Native Speakers," *Alsinatuna* 10, no. 1 (2024), https://e-journal.uingusdur.ac.id/alsinatuna/article/view/7952.

[8] Shengpeng Ji et al., "TextrolSpeech: A Text Style Control Speech Corpus with Codec Language Text-to-Speech Models," *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 14, 2024, 10301–5, https://doi.org/10.1109/ICASSP48485.2024.10445879; Matan P and P. Velvizhy, "A Comprehensive Review of Expressive Text-To-Speech Systems and Its Advancements and Challenges," *2025 International Conference on Inventive Computation Technologies (ICICT)*, April 23, 2025, 49–54, https://doi.org/10.1109/ICICT64420.2025.11004848.

Arabic serves as a key language in pesantren and Islamic higher education in Indonesia as a gateway to accessing Islamic religious knowledge and classical scholarly traditions.[9] Listening skill development, which is vital for mastering the language and seeking Islamic teachings, requires additional attention.[10] Limited exposure to authentic audio recordings constrains students' proficiency development. ElevenLabs text-to-speech (TTS) technology offers a potential solution by generating extensive Arabic audio from written sources.[11] The tool allows teachers to adapt materials to match students' listening levels, thereby facilitating input towards sustainability in pesantren secular-subject education.[12]

In recent years, the integration of artificial intelligence in language education has been increasingly conceptualized not merely as a technological enhancement but as a paradigm shift in how linguistic input is mediated, perceived, and internalized by learners.[13] AI-driven auditory technologies, particularly advanced Text-to-Speech (TTS) systems, have been shown to reshape the cognitive ecology of listening by offering controllable, repeatable, and affectively rich input that traditional classroom narration often cannot sustain.[14] Scholars in second language acquisition argue that listening comprehension is highly sensitive to input quality, prosodic naturalness, and emotional salience, factors that directly influence attention, memory consolidation, and meaning-making processes.[15] Within this framework, AI-based TTS emerges as a mediating tool that operationalizes core SLA theories such as comprehensible input, cognitive load management, and multimodal processing into practical instructional resources.[16] Consequently, the study of AI-generated speech is no longer confined to engineering or computational linguistics

[9] Nur Hanifansyah et al., "Religious Drama Controversy: The Impact of Bidaah on Islamic Pedagogy and Media Literacy," *MIQOT: Jurnal Ilmu-Ilmu Keislaman* 49, no. 2 (2025): 314, https://doi.org/10.30821/miqot.v49i2.1407.

[10] Segaf Baharun et al., "Creative Arabic Learning through Student-Made Storytelling: A Constructivist Approach in Malaysian Islamic Schools," *Arabi : Journal of Arabic Studies* 10, no. 2 (2025): 149–61, https://doi.org/10.24865/ajas.v10i2.1008.

[11] Ferdinand De Saussure, *Writings In General Linguistics* (Oxford University PressOxford, 2006), https://doi.org/10.1093/oso/9780199261444.001.0001.

[12] Abdelrahman Mohamed et al., "Self-Supervised Speech Representation Learning: A Review," *IEEE Journal of Selected Topics in Signal Processing* 16, no. 6 (2022): 1179–210, https://doi.org/10.1109/JSTSP.2022.3207050.

[13] Ruojun Zhong and Yong Zhao, "Education Paradigm Shifts in the Age of AI: A Spatiotemporal Analysis of Learning," *ECNU Review of Education* 8, no. 2 (2025): 319–42, https://doi.org/10.1177/20965311251315204.

[14] Mohamed Sayed Abdellatif et al., "I Am All Ears: Listening Exams with AI and Its Traces on Foreign Language Learners' Mindsets, Self-Competence, Resilience, and Listening Improvement," *Language Testing in Asia* 14, no. 1 (2024): 54, https://doi.org/10.1186/s40468-024-00329-6.

[15] V. Madhusudhana Reddy et al., "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning," *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, July 19, 2023, 657–66, https://doi.org/10.1109/ICECAA58104.2023.10212222.

[16] Sumukh Sirmokadam, "Speech To Text for Data Entry—Opportunities and Challenges," in *Data Management, Analytics and Innovation*, vol. 137, Lecture Notes on Data Engineering and Communications Technologies (Springer Nature Singapore, 2023), https://doi.org/10.1007/978-981-19-2600-6_25.

but has become a central concern in applied linguistics, educational psychology, and digital pedagogy.[17]

Recent studies have underscored the growing integration of technology and artificial intelligence in enhancing listening comprehension and auditory learning. Mulyadi et al. demonstrated that technology-enhanced task-based instruction significantly improved learners' listening and speaking performance, emphasizing the value of digital tasks in facilitating active engagement.[18] Similarly, Tilwani et al. confirmed that authentic digital materials, such as TED Talks, provide more effective listening input than conventional media, fostering motivation and comprehension.[19] Keelor et al. further established that text-to-speech (TTS) technology supports learners with language difficulties by improving comprehension and reducing cognitive load.[20] Torres et al. provided empirical evidence of the high acoustic fidelity of ElevenLabs-generated voices, showing near-human spectral quality suitable for linguistic applications.[21] Complementing this, Ada et al. explored the paralinguistic dimensions of AI-generated voices, revealing that ElevenLabs can reproduce human-like prosody, pacing, and emotional tone, thereby expanding the expressive potential of AI in language learning.[22] Collectively, these studies confirm that technology particularly AI-based voice synthesis can enhance listening comprehension by providing natural, authentic, and affectively engaging auditory input.

Despite these advancements, prior research has predominantly focused on English as a foreign language or general speech synthesis contexts, leaving a significant gap in understanding the pedagogical application of AI text-to-speech technologies in Arabic language education, especially within Islamic boarding schools (pesantren). None of the existing studies have investigated how ElevenLabs can be pedagogically adapted to foster Arabic listening comprehension within a spiritually oriented and ethically grounded environment. Furthermore, while earlier works examined acoustic authenticity and paralinguistic expressiveness, little is known about how such features influence learners' cognitive processing,

[17] Jonathan M. Golding et al., "Generative AI and College Students: Use and Perceptions," *Teaching of Psychology* 52, no. 3 (2025): 369–80, https://doi.org/10.1177/00986283241280350.

[18] Dodi Mulyadi et al., "Effects of Technology Enhanced Task-Based Language Teaching on Learners' Listening Comprehension and Speaking Performance," *International Journal of Instruction* 14, no. 3 (2021): 717–36, https://doi.org/10.29333/iji.2021.14342a.

[19] Shouket Ahmad Tilwani et al., "The Impact of Using TED Talks as a Learning Instrument on Enhancing Indonesian EFL Learners' Listening Skill," *Education Research International* 2022 (March 2022): 1–9, https://doi.org/10.1155/2022/8036363.

[20] Jennifer L. Keelor et al., "Impact of Text-to-Speech Features on the Reading Comprehension of Children with Reading and Language Difficulties," *Annals of Dyslexia* 73, no. 3 (2023): 469–86, https://doi.org/10.1007/s11881-023-00281-9.

[21] Hailun Lian et al., "A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face," *Entropy* 25, no. 10 (2023): 1440, https://doi.org/10.3390/e25101440.

[22] Ada Ada Ada et al., "Cultures of the AI Paralinguistic in Voice Cloning Tools," *Designing Interactive Systems Conference*, July 2024, 249–52, https://doi.org/10.1145/3656156.3663708; Xinfa Zhu et al., "METTS: Multilingual Emotional Text-to-Speech by Cross-Speaker and Cross-Lingual Emotion Transfer," *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32 (2024): 1506–18, https://doi.org/10.1109/TASLP.2024.3363444.

motivation, and spiritual engagement in non-native Arabic contexts. Therefore, this study addresses these gaps by examining the effectiveness of ElevenLabs TTS in enhancing Arabic listening skills in pesantren settings, bridging technological innovation with traditional Islamic pedagogy and contributing to the emerging discourse on AI-assisted Arabic language learning.

Specifically, this study is guided by two main research questions: To what extent does ElevenLabs Text-to-Speech technology improve students' listening comprehension in Arabic learning at pesantren? How do students perceive the use of ElevenLabs TTS in terms of clarity, motivation, and engagement compared to traditional listening methods?

The scope of this research is limited to intermediate-level Arabic learners at Pondok Pesantren Darullughah Wadda'wah, where ElevenLabs-generated materials were integrated into the existing listening curriculum over six weeks.[23] The study does not address speaking or pronunciation training directly but focuses on listening comprehension and auditory perception. Moreover, while ElevenLabs supports multiple languages, this study is restricted to Modern Standard Arabic (fusha) audio to maintain linguistic consistency and pedagogical relevance.[24]

The significance of this research lies in its contribution to both pedagogical innovation and technological integration in Islamic education. Pedagogically, it provides empirical evidence that AI voice synthesis can complement human instruction by offering consistent, expressive, and accessible Arabic input. Technologically, it pioneers the contextual adaptation of ElevenLabs for moral and spiritual learning settings, bridging the gap between traditional pesantren pedagogy and modern digital tools. By situating the study within the intersection of AI, Arabic education, and Islamic pedagogy, this research aspires to demonstrate that intelligent audio technologies can humanize – not replace – the act of listening as both a linguistic and spiritual discipline.

Method

This study employed a mixed-methods design combining quantitative and qualitative approaches to examine the effectiveness of ElevenLabs Text-to-Speech (TTS) technology in enhancing Arabic listening skills among pesantren students.[25] The design was grounded in Krashen's Input Hypothesis, Mayer's Cognitive Theory of Multimedia Learning, and Paivio's Dual Coding Theory, emphasizing the importance of comprehensible input, multimedia integration, and dual cognitive processing in second language learning.[26] By merging measurable learning

[23] Segaf Baharun et al., "The I'rab Method of Habib Hasan Baharun: Impact on Arabic Grammar Instruction," *Al-Muhawaroh: Jurnal Pendidikan Bahasa Arab* 1, no. 1 (2025): 23–35, https://doi.org/10.38073/almuhawaroh.v1i1.2636.

[24] Samah M. Alzanin et al., "Short Text Classification for Arabic Social Media Tweets," *Journal of King Saud University - Computer and Information Sciences* 34, no. 9 (2022): 6595–604, https://doi.org/10.1016/j.jksuci.2022.03.020.

[25] John W. Creswell and J. David Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Los Angeles (SAGE Publications, 2020).

[26] Stephen D. Krashen, "Acquiring A Second Language," *World Englishes* 1, no. 3 (1982): 97–101, https://doi.org/10.1111/j.1467-971X.1982.tb00476.x; Alessio Di Paolo, "The Relation Between Mayer's

outcomes with interpretive classroom data, this approach allowed a holistic evaluation of students' linguistic and affective development.

The study adopted a quasi-experimental pre-test–post-test design, complemented by interviews and observations.[27] It was conducted at Pondok Pesantren Darullughah Wadda'wah in Pasuruan, East Java, an institution chosen for its strong Arabic learning tradition and its emerging use of digital learning tools. Sixty intermediate-level students aged 17–21 were selected and divided equally into control and experimental groups. The experimental group used Arabic listening materials generated through ElevenLabs TTS, while the control group used teacher-narrated audio.

The selection of participants followed a purposive sampling procedure, based on institutional placement records indicating intermediate-level Arabic proficiency. From three existing parallel classes at the same academic level, two classes with comparable mean listening scores in the previous semester were chosen to ensure baseline equivalence. One intact class was assigned as the experimental group and the other as the control group to maintain natural classroom conditions and avoid disruption of the pesantren's instructional structure. Prior to the intervention, an initial pre-test was administered to confirm that no statistically significant differences existed between the two groups' listening proficiency levels. This procedure ensured internal validity while maintaining ecological validity within the quasi-experimental framework.

Both primary and secondary data were collected. Primary data included pre–post listening test scores, student interviews, and classroom observation notes, while secondary data comprised institutional documents and prior research on AI speech synthesis in education.[28] Instruments used were a listening comprehension test based on CEFR B1–B2 proficiency levels, an interview guide to assess learner perceptions, and an observation checklist to record classroom dynamics and motivation.[29] The research lasted six weeks, with two listening sessions per week involving pre-listening discussion, ElevenLabs-based listening activities, and post-listening comprehension tasks.

Multimedia Theory and Berthoz's Simplexity Paradigm for Inclusive Education," in *Advanced Research in Technologies, Information, Innovation and Sustainability*, vol. 1936, Communications in Computer and Information Science (Springer Nature Switzerland, 2024), https://doi.org/10.1007/978-3-031-48855-9_24; Mark Sadoski and Allan Paivio, *Imagery and Text: A Dual Coding Theory of Reading and Writing*, 0 ed. (Routledge, 2012), https://doi.org/10.4324/9781410605276.

[27] Ferian Fauzi Abdulloh et al., "Observation of Imbalance Tracer Study Data for Graduates Employability Prediction in Indonesia," *International Journal of Advanced Computer Science and Applications* 13, no. 8 (2022), https://doi.org/10.14569/IJACSA.2022.0130820.

[28] Nicole M. Deterding and Mary C. Waters, "Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach," *Sociological Methods & Research* 50, no. 2 (2021): 708–39, https://doi.org/10.1177/0049124118799377; Charlotte Dignath and Marcel V. J. Veenman, "The Role of Direct Strategy Instruction and Indirect Activation of Self-Regulated Learning—Evidence from Classroom Observation Studies," *Educational Psychology Review* 33, no. 2 (2021): 489–533, https://doi.org/10.1007/s10648-020-09534-0.

[29] Omolola A. Adeoye-Olatunde and Nicole L. Olenik, "Research and Scholarly Methods: Semi-structured Interviews," *JACCP: Journal of the American College of Clinical Pharmacy* 4, no. 10 (2021): 1358–67, https://doi.org/10.1002/jac5.1441.

Data were analyzed quantitatively using SPSS for descriptive statistics and paired sample t-tests, while qualitative data were processed thematically using NVivo 14 through open coding and theme categorization. Emerging themes such as clarity, motivation, and spiritual engagement were triangulated across instruments to ensure validity. Ethical standards were maintained by obtaining informed consent, ensuring confidentiality, and using all AI-generated materials solely for educational purposes in accordance with pesantren values.

This methodological framework combined empirical rigor with pedagogical sensitivity, aligning cognitive-linguistic theories with the ethical and spiritual atmosphere of pesantren education to assess the true impact of ElevenLabs TTS on Arabic listening development.

Result and Discussion

Effect of ElevenLabs TTS on Students' Listening Comprehension

To examine the impact of ElevenLabs Text-to-Speech technology on Arabic listening development, a quantitative analysis was conducted to compare students' listening performance before and after the six-week instructional intervention. The analysis focused on identifying whether the use of AI-generated Arabic audio could significantly enhance students' comprehension accuracy, phonological recognition, and processing efficiency compared with conventional teacher-narrated listening materials.

Quantitative analysis using paired sample t-tests revealed a significant improvement in students' Arabic listening comprehension after six weeks of exposure to ElevenLabs-generated materials. The experimental group's mean score increased from 63.2 to 82.7, while the control group improved modestly from 64.5 to 71.4. The p-value of 0.001 ($p < 0.05$) indicated a statistically significant difference between the two groups, confirming that the use of ElevenLabs TTS had a substantial positive effect on students' comprehension accuracy, word recognition, and phonological awareness. Students exposed to the AI-generated audio displayed a better understanding of connected speech and intonation patterns, as well as faster response times in comprehension tasks.

To provide clearer statistical interpretation, the magnitude of improvement in the experimental group corresponds to a large practical effect, as indicated by the substantial mean gain (+19.5 points) compared to the control group (+6.9 points). While statistical significance ($p < 0.05$) confirms that the difference was unlikely due to chance, the size of the gain demonstrates meaningful pedagogical impact. In practical terms, students in the experimental group progressed from a moderate level of comprehension to a high-intermediate performance range within six weeks, reflecting not only improved accuracy but also enhanced processing efficiency. This suggests that the observed improvement represents substantial learning advancement rather than a marginal score fluctuation.

These findings suggest that ElevenLabs TTS offers high-quality auditory input that aligns with Krashen's Input Hypothesis, where learners acquire language effectively when exposed to comprehensible and engaging input. The natural prosody and emotional tone embedded in ElevenLabs voices helped students

process meaning more intuitively and reduce cognitive overload, consistent with Mayer's Cognitive Theory of Multimedia Learning. Compared with traditional teacher narration, ElevenLabs provided a level of voice consistency and expressive clarity, which enhanced the precision of learners' listening comprehension.

This result supports previous findings by Zhang and MacWhinney, who demonstrated that AI-generated audio can significantly improve L2 listening comprehension through enhanced speech naturalness and pacing control. [30] Similarly, Vogel et al. noted that emotional realism in TTS voices can foster stronger auditory engagement, particularly among learners who struggle with sustained attention. [31] In the pesantren context, where listening is both a linguistic and spiritual exercise, ElevenLabs TTS bridged this dual function effectively, maintaining clarity and engagement without compromising the authenticity of Arabic expression.

Students' Perceptions and Learning Experience with ElevenLabs TTS

Qualitative data from interviews and classroom observations further reinforced the quantitative outcomes. Students consistently described ElevenLabs-generated voices as "clear," "authentic," and "emotionally expressive." Several learners expressed that they felt "closer to real Arabic speakers" when listening to the AI audio compared to traditional teacher recordings. One student remarked: *"When I listen to the ElevenLabs voice, it feels like a native Arabic teacher speaking. The tone is natural and helps me focus on meaning rather than just words."* Another student reflected on how the technology reduced anxiety during listening tasks: *"Sometimes when the teacher speaks fast, I feel nervous and cannot follow. But with ElevenLabs, I can replay and focus again. It helps me understand without pressure."*

Teachers also noted that the students appeared more attentive during the sessions and more confident in subsequent speaking activities, suggesting that improved listening comprehension contributed indirectly to better oral fluency. Observation data revealed that students often took personal notes on pronunciation and prosody, showing higher metacognitive awareness compared to previous sessions with traditional materials.

These qualitative insights illustrate that ElevenLabs TTS not only supports linguistic comprehension but also enhances affective engagement and learner autonomy. The ability to adjust playback and maintain consistency of tone allowed students to internalize Arabic sounds in a relaxed, self-paced manner—an advantage rarely found in conventional classroom listening.

[30] Yanhui Zhang and Brian MacWhinney, "Using Diagnostic Feedback to Enhance the Development of Phonetic Knowledge of an L2: A CALL Design Based on the Unified Competition Model and the Implementation with the Pinyin Tutor," *Language Testing in Asia* 13, no. 1 (2023): 35, https://doi.org/10.1186/s40468-023-00232-6.

[31] Adam P. Vogel et al., "Optimizing Communication in Ataxia: A Multifaceted Approach to Alternative and Augmentative Communication (AAC)," *The Cerebellum* 23, no. 5 (2024): 2142–51, https://doi.org/10.1007/s12311-024-01675-0.

Table 1. Descriptive Statistics of Students' Arabic Listening Scores

Group	N	Pre-Test Mean	Post-Test Mean	Mean Gain
Experimental Group (ElevenLabs TTS)	30	63.2	82.7	+19.5
Control Group (Teacher-Narrated Audio)	30	64.5	71.4	+6.9

Table 1 shows that both groups experienced improvement in Arabic listening comprehension; however, the experimental group exposed to ElevenLabs TTS demonstrated a substantially higher mean gain compared with the control group. This indicates that AI-generated auditory input contributed more effectively to listening development than conventional teacher-narrated audio.

Beyond the numerical comparison, the data indicate a clear divergence in learning trajectories between the two groups. The experimental group's increase of 19.5 points represents a substantial developmental shift over a relatively short six-week intervention, suggesting accelerated listening acquisition rather than gradual improvement. In contrast, the control group's gain of 6.9 points reflects incremental progress typically expected from regular instructional exposure. This contrast implies that the AI-generated auditory input did not merely supplement existing pedagogy but functioned as a catalytic factor in enhancing phonological processing and comprehension efficiency. The magnitude of the difference suggests that structured exposure to consistent, native-like prosodic patterns may have strengthened students' auditory discrimination and reduced processing strain during listening tasks.

Table 2. Paired Sample t-Test Results for Listening Comprehension

Group	t-value	Sig. (p-value)	Result
Experimental Group	4.87	0.001	Significant
Control Group	1.92	0.064	Not Significant

As presented in Table 2, the experimental group showed a statistically significant improvement in listening comprehension scores ($p < 0.05$), whereas the control group did not reach statistical significance. This confirms that the observed improvement in the experimental group was not due to chance and can be attributed to the use of ElevenLabs TTS technology.

Table 3. Qualitative Themes from Student Interviews

Theme	Description	Representative Student Response
Clarity of Audio	Improved sound articulation and phoneme distinction	*"The voice is very clear and easy to understand."*
Authenticity	Native-like pronunciation and prosody	*"It feels like listening to a real Arabic speaker."*
Emotional Engagement	Expressive tone enhancing motivation	*"The voice makes learning more enjoyable."*
Learning Autonomy	Ability to replay and self-regulate learning pace	*"I can repeat the audio without feeling pressured."*

The findings confirm that AI-based auditory tools like ElevenLabs can effectively enhance Arabic listening skills in pesantren contexts by providing rich, emotionally nuanced, and cognitively manageable input. The improvement in comprehension scores validates Krashen's claim that sustained exposure to comprehensible input facilitates natural acquisition. The qualitative results further affirm Mayer's multimedia learning principles, as students engaged both auditory and cognitive channels more efficiently when listening to expressive, human-like AI voices.

The convergence of quantitative improvement and qualitative insight underscores the pedagogical, cognitive, and spiritual effectiveness of ElevenLabs TTS technology. It not only enhances Arabic listening comprehension but also transforms the classroom atmosphere, making learning more autonomous, emotionally engaging, and spiritually meaningful.

The findings of this study align with and meaningfully extend a growing body of scholarship affirming that AI-mediated auditory tools can substantially transform language learning outcomes. The significant improvement recorded in the experimental group, whose listening scores rose from 63.2 to 82.7 within six weeks, confirms that ElevenLabs TTS delivers a form of input that is not merely technically superior, but pedagogically potent. This result is consistent with Krashen's Input Hypothesis, which posits that learners acquire language most effectively when exposed to comprehensible, meaningful, and emotionally accessible input that slightly exceeds their current proficiency level.[32]

From the perspective of cognitive science, the gains observed in the experimental group can be further interpreted through Roussel et al.'s cognitive load framework, which demonstrated that when learners access audio content with reduced processing strain such as consistent pacing and natural prosody both linguistic and content acquisition improve significantly. [33] ElevenLabs TTS, by offering stable, repeatable auditory input free from the unpredictability of live human narration, effectively reduces extraneous cognitive load and directs learners' working memory toward deeper phonological mapping and semantic processing. This interpretation resonates with findings by Keelor et al., who established that TTS technology helps learners with language processing difficulties by enhancing comprehension and lowering cognitive demands.[34] The current study extends this insight to an Arabic L2 population in pesantren contexts, where phonological complexity particularly emphatic consonants, pharyngeals, and rhythmic patterns poses heightened perceptual demands on non-native ears.

The affective dimension of this study deserves particular theoretical attention. Students described the ElevenLabs voice as emotionally expressive and motivating qualities that align directly with what Xiao identified as "flow experience" in AI-

[32] Krashen, "Acquiring A Second Language."

[33] Stéphanie Roussel et al., "The Advantages of Listening to Academic Content in a Second Language May Be Outweighed by Disadvantages: A Cognitive Load Theory Approach," *British Journal of Educational Psychology* 92, no. 2 (2022): 627–44, https://doi.org/10.1111/bjep.12468.

[34] Keelor et al., "Impact of Text-to-Speech Features on the Reading Comprehension of Children with Reading and Language Difficulties."

driven listening interventions. In a randomized controlled trial, Xiao found that AI-driven speech recognition not only improved EFL learners' listening comprehension but also reduced listening anxiety and sustained engagement over time.[35] The parallels are striking: in both studies, AI-mediated input creates a psychologically safe environment where learners disengage from performance-based anxiety and re-engage with meaning-making. This aligns with earlier findings by Fathi et al., who reported that AI-mediated interaction significantly improved EFL learners' communicative willingness and reduced fear of real-time language use.[36] In the pesantren context, where students are simultaneously managing high-stakes religious and linguistic learning, the anxiety-reduction afforded by ElevenLabs TTS represents not just a technical convenience but also profound pedagogical relief.

The question of learner autonomy, one of the most consistent themes in the qualitative data, also invites comparison with recent theoretical work on self-regulated learning (SRL) in AI-enhanced environments. Learners in the experimental group consistently reported the ability to replay audio, control the listening pace, and monitor their own comprehension without depending on teacher interruption. This pattern maps closely onto findings reported in a systematic review by Mohebbi, which found that AI tools significantly foster learner independence and metacognitive strategy use by enabling self-paced, self-monitored learning trajectories.[37] Critically, in the pesantren environment, where collective learning norms have traditionally constrained individual agency, such autonomy constitutes a meaningful pedagogical innovation, enabling students to internalize Arabic phonology on their own cognitive timeline without disrupting communal learning rhythms.

The deep-learning mechanisms underlying ElevenLabs' prosodic authenticity also warrant scholarly attention. Khairnar and Velvizhy, in a comprehensive review of expressive TTS systems, noted that contemporary neural TTS architectures have achieved near-human prosodic fidelity by modeling emotional contour, speech rhythm, and phonemic boundary transitions simultaneously.[38] This technical evolution directly explains why students in the current study perceived the AI voice as "native-like" and "authentic" qualities that stimulate phonological sensitivity and strengthen auditory discrimination, skills foundational to Arabic listening competence. Furthermore, Chung demonstrated that prosodic awareness, defined as sensitivity to pitch, rhythm, and amplitude, is a significant predictor of second

[35] Yanling Xiao, "The Impact of AI-Driven Speech Recognition on EFL Listening Comprehension, Flow Experience, and Anxiety: A Randomized Controlled Trial," *Humanities and Social Sciences Communications* 12, no. 1 (2025): 425, https://doi.org/10.1057/s41599-025-04672-8.

[36] Jalil Fathi et al., "Improving EFL Learners' Speaking Skills and Willingness to Communicate via Artificial Intelligence-Mediated Interactions," *System* 121 (April 2024): 103254, https://doi.org/10.1016/j.system.2024.103254.

[37] Ahmadreza Mohebbi, "Enabling Learner Independence and Self-Regulation in Language Education Using AI Tools: A Systematic Review," *Cogent Education* 12, no. 1 (2025): 2433814, https://doi.org/10.1080/2331186X.2024.2433814.

[38] P and Velvizhy, "A Comprehensive Review of Expressive Text-To-Speech Systems and Its Advancements and Challenges."

language word learning, even independently of general proficiency levels.[39] The ElevenLabs voice, by providing consistent, prosodically rich Arabic input, may therefore serve as a training stimulus for the very auditory-perceptual mechanisms that underpin long-term listening acquisition.

Beyond cognitive and affective gains, this study positions ElevenLabs TTS within the unique ethical and spiritual ecology of pesantren education, a dimension absent from prior AI-language research conducted in secular Western contexts. The concept of adab al-sam' (ethical listening) in Islamic pedagogy implies that the act of listening is not merely receptive but spiritually formative: a discipline of attention, humility, and meaning-seeking. Rather than displacing this tradition, ElevenLabs appears to reinforce it. Students who engaged with the AI audio were observed to take careful notes on pronunciation and prosody, demonstrating a metacognitive attentiveness that mirrors the spirit of istima' as a sacred act of engagement. This finding resonates with Chen and Lehto's observation that AI-synthesized voices, when acoustically designed with emotional naturalismcan enhance the depth of listening experience in educational contexts, suggesting that technology need not be ethically neutral but can be instrumentalized to deepen reflective learning.

Taken together, these multidimensional findings construct a compelling argument: ElevenLabs TTS is not simply a more convenient substitute for teacher narration. It is a cognitively facilitative, affectively enriching, and spiritually compatible tool for Arabic language learning in pesantren – one that simultaneously addresses the phonological, psycholinguistic, and motivational challenges specific to non-native Arabic learners in Islamic educational contexts. Future research should investigate its long-term effects across multiple institutions and across the full spectrum of Arabic language skills, while also interrogating questions of learner dependency, technological equity, and the limits of AI authenticity in culturally distinct pedagogical settings.

Conclusion

This study demonstrates that ElevenLabs Text-to-Speech (TTS) technology significantly enhances students' Arabic listening comprehension within the unique context of pesantren education. The integration of AI-generated audio improved learners' accuracy, phonological awareness, and motivation by providing consistent, expressive, and authentic Arabic input. Quantitative results confirmed a substantial increase in listening performance, while qualitative insights revealed higher engagement, reduced anxiety, and stronger learner autonomy. These outcomes validate Krashen's Input Hypothesis and Mayer's Cognitive Theory of Multimedia Learning, affirming that emotionally rich and comprehensible auditory input promotes deeper linguistic processing. ElevenLabs thus proves effective not only as a technological innovation but also as a pedagogical bridge that aligns modern AI learning tools with the spiritual and ethical values of Islamic education.

[39] Wei-Lun Chung, "General Auditory Processing, Mandarin L1 Prosodic and Phonological Awareness, and English L2 Word Learning," *International Review of Applied Linguistics in Language Teaching* 63, no. 3 (2025): 1895–914, https://doi.org/10.1515/iral-2023-0168.

Specifically, the experimental group's listening scores increased from a pre-test mean of 63.2 to a post-test mean of 82.7, reflecting a mean gain of 19.5 points over six weeks. In comparison, the control group improved from 64.5 to 71.4, with a mean gain of 6.9 points. This difference, supported by a statistically significant result ($p = 0.001$), indicates that students exposed to ElevenLabs-generated audio achieved nearly three times greater improvement than those receiving conventional teacher-narrated input. These findings demonstrate not only statistical significance but also meaningful pedagogical impact within a relatively short instructional period.

Despite its success, the study acknowledges certain limitations, including the relatively short intervention period and the focus on a single institution, which may constrain generalizability. Future research could explore long-term applications of ElevenLabs in other Arabic skills – particularly speaking and pronunciation – or extend its use to diverse pesantren settings with varying pedagogical traditions. Practically, this research offers valuable implications for educators: integrating AI-based TTS tools like ElevenLabs can complement traditional teaching by providing accessible, high-quality listening materials that promote independent learning. When thoughtfully contextualized, such innovations can modernize Arabic instruction while preserving the human and spiritual dimensions that define pesantren education.

Acknowledgment

The authors would like to express their deepest gratitude to Universitas Sains Al-Qur'an, Universitas Islam Internasional Darullughah Wadda'wah, and International Islamic University Malaysia for their academic and institutional support. Special thanks are extended to the teachers and students of Pondok Pesantren Darullughah Wadda'wah, whose participation and enthusiasm made this research possible. The authors also acknowledge the assistance of the Arabic language education department for facilitating the implementation of the AI-based listening sessions using ElevenLabs technology.

Author Contribution Statement

AS conceptualized the research design, supervised the data collection process, and drafted the main manuscript. MS contributed to the development of the theoretical framework, literature review, and validation of the listening comprehension instruments. NS performed the data analysis, assisted in writing the discussion section, and refined the English language of the final manuscript. All authors read and approved the final version of the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AI Writing Statement

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) only for language editing and grammatical improvement. All scientific content, analysis, and interpretations were developed by the authors. The authors carefully reviewed and edited the AI-assisted output and take full responsibility for the final content of this manuscript.

References

Abdellatif, Mohamed Sayed, Mohammed A. Alshehri, Hamoud A. Alshehri, Waheed Elsayed Hafez, Mona G. Gafar, and Ali Lamouchi. "I Am All Ears: Listening Exams with AI and Its Traces on Foreign Language Learners' Mindsets, Self-Competence, Resilience, and Listening Improvement." *Language Testing in Asia* 14, no. 1 (2024): 54. https://doi.org/10.1186/s40468-024-00329-6.

Abdulloh, Ferian Fauzi, Majid Rahardi, Afrig Aminuddin, Sharazita Dyah Anggita, and Arfan Yoga Aji Nugraha. "Observation of Imbalance Tracer Study Data for Graduates Employability Prediction in Indonesia." *International Journal of Advanced Computer Science and Applications* 13, no. 8 (2022). https://doi.org/10.14569/IJACSA.2022.0130820.

Ada, Ada Ada, Stina Hasse Jørgensen, and Jonas Fritsch. "Cultures of the AI Paralinguistic in Voice Cloning Tools." *Designing Interactive Systems Conference*, July 2024, 249–52. https://doi.org/10.1145/3656156.3663708.

Adeoye-Olatunde, Omolola A., and Nicole L. Olenik. "Research and Scholarly Methods: Semi-structured Interviews." *JACCP: Journal of the American College of Clinical Pharmacy* 4, no. 10 (2021): 1358–67. https://doi.org/10.1002/jac5.1441.

Ahmad, Maysa, Ahmad S. Haider, and Hadeel Saed. "Assessing AI-Driven Dubbing Websites: Reactions of Arabic Native Speakers to AI-Dubbed English Videos in Arabic." *Research Journal in Advanced Humanities* 6, no. 1 (2025). https://royalliteglobal.com/advanced-humanities/article/view/1963.

Alzanin, Samah M., Aqil M. Azmi, and Hatim A. Aboalsamh. "Short Text Classification for Arabic Social Media Tweets." *Journal of King Saud University - Computer and Information Sciences* 34, no. 9 (2022): 6595–604. https://doi.org/10.1016/j.jksuci.2022.03.020.

Arisandi, Yusuf, and Moh. Tohiri Habib. "Optimizing YouTube for Interactive Arabic Learning in Pesantren: Effective Content Creation Strategies." *International Journal of Arabic Language Teaching* 7, no. 02 (2025): 239–54. https://doi.org/10.32332/ijalt.v7i02.10363.

Baharun, Segaf, Nur Hanifansyah, and Aufa Hanin Salsabil. "Creative Arabic Learning through Student-Made Storytelling: A Constructivist Approach in Malaysian

Islamic Schools." *Arabi : Journal of Arabic Studies* 10, no. 2 (2025): 149–61. https://doi.org/10.24865/ajas.v10i2.1008.

Baharun, Segaf, Muhamad Solehudin, Masnun, and Syarif Muhammad Syaheed. "The I'rab Method of Habib Hasan Baharun: Impact on Arabic Grammar Instruction." *Al-Muhawaroh: Jurnal Pendidikan Bahasa Arab* 1, no. 1 (2025): 23–35. https://doi.org/10.38073/almuhawaroh.v1i1.2636.

Bsharat-Maalouf, Dana, Tamar Degani, and Hanin Karawani. "The Involvement of Listening Effort in Explaining Bilingual Listening Under Adverse Listening Conditions." *Trends in Hearing* 27 (January 2023): 23312165231205107. https://doi.org/10.1177/23312165231205107.

Chen, Jun, and Xinran Lehto. "The Impact of Sound Design with AI Synthetic Voices on the Listening Experience in Audio Tour Guides." *Information Technology & Tourism* 27, no. 4 (2025): 1081–109. https://doi.org/10.1007/s40558-025-00332-4.

Chung, Wei-Lun. "General Auditory Processing, Mandarin L1 Prosodic and Phonological Awareness, and English L2 Word Learning." *International Review of Applied Linguistics in Language Teaching* 63, no. 3 (2025): 1895–914. https://doi.org/10.1515/iral-2023-0168.

Creswell, John W., and J. David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Los Angeles. SAGE Publications, 2020.

Deterding, Nicole M., and Mary C. Waters. "Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach." *Sociological Methods & Research* 50, no. 2 (2021): 708–39. https://doi.org/10.1177/0049124118799377.

Di Paolo, Alessio. "The Relation Between Mayer's Multimedia Theory and Berthoz's Simplexity Paradigm for Inclusive Education." In *Advanced Research in Technologies, Information, Innovation and Sustainability*, vol. 1936. Communications in Computer and Information Science. Springer Nature Switzerland, 2024. https://doi.org/10.1007/978-3-031-48855-9_24.

Dignath, Charlotte, and Marcel V. J. Veenman. "The Role of Direct Strategy Instruction and Indirect Activation of Self-Regulated Learning—Evidence from Classroom Observation Studies." *Educational Psychology Review* 33, no. 2 (2021): 489–533. https://doi.org/10.1007/s10648-020-09534-0.

Fathi, Jalil, Masoud Rahimi, and Ali Derakhshan. "Improving EFL Learners' Speaking Skills and Willingness to Communicate via Artificial Intelligence-Mediated Interactions." *System* 121 (April 2024): 103254. https://doi.org/10.1016/j.system.2024.103254.

Golding, Jonathan M., Anne Lippert, Jeffrey S. Neuschatz, Ilyssa Salomon, and Kelly Burke. "Generative AI and College Students: Use and Perceptions." *Teaching of Psychology* 52, no. 3 (2025): 369–80. https://doi.org/10.1177/00986283241280350.

Hanifansyah, Nur, Ahmad Arifin, Zulpina Zulpina, Menik Mahmudah, and Syarif Muhammad Syaheed. "Religious Drama Controversy: The Impact of Bidaah on Islamic Pedagogy and Media Literacy." *MIQOT: Jurnal Ilmu-Ilmu Keislaman* 49, no. 2 (2025): 314. https://doi.org/10.30821/miqot.v49i2.1407.

Ji, Shengpeng, Jialong Zuo, Minghui Fang, et al. "TextrolSpeech: A Text Style Control Speech Corpus with Codec Language Text-to-Speech Models." *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 14, 2024, 10301–5. https://doi.org/10.1109/ICASSP48485.2024.10445879.

Keelor, Jennifer L., Nancy A. Creaghead, Noah H. Silbert, Allison D. Breit, and Tzipi Horowitz-Kraus. "Impact of Text-to-Speech Features on the Reading Comprehension of Children with Reading and Language Difficulties." *Annals of Dyslexia* 73, no. 3 (2023): 469–86. https://doi.org/10.1007/s11881-023-00281-9.

Krashen, Stephen D. "Acquiring A Second Language." *World Englishes* 1, no. 3 (1982): 97–101. https://doi.org/10.1111/j.1467-971X.1982.tb00476.x.

Kumar, Yogesh, Apeksha Koul, and Chamkaur Singh. "A Deep Learning Approaches in Text-to-Speech System: A Systematic Review and Recent Research Perspective." *Multimedia Tools and Applications* 82, no. 10 (2023): 15171–97. https://doi.org/10.1007/s11042-022-13943-4.

Lian, Hailun, Cheng Lu, Sunan Li, Yan Zhao, Chuangao Tang, and Yuan Zong. "A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face." *Entropy* 25, no. 10 (2023): 1440. https://doi.org/10.3390/e25101440.

Ma, Ziyang, Wen Wu, Zhisheng Zheng, et al. "Leveraging Speech PTM, Text LLM, And Emotional TTS For Speech Emotion Recognition." *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 14, 2024, 11146–50. https://doi.org/10.1109/ICASSP48485.2024.10445906.

Mohamed, Abdelrahman, Hung-yi Lee, Lasse Borgholt, et al. "Self-Supervised Speech Representation Learning: A Review." *IEEE Journal of Selected Topics in Signal Processing* 16, no. 6 (2022): 1179–210. https://doi.org/10.1109/JSTSP.2022.3207050.

Mohebbi, Ahmadreza. "Enabling Learner Independence and Self-Regulation in Language Education Using AI Tools: A Systematic Review." *Cogent Education* 12, no. 1 (2025): 2433814. https://doi.org/10.1080/2331186X.2024.2433814.

Mulyadi, Dodi, Testiana Deni Wijayatiningsih, Charanjit Kaur Swaran Singh, and Entika Fani Prastikawati. "Effects of Technology Enhanced Task-Based Language Teaching on Learners' Listening Comprehension and Speaking Performance." *International Journal of Instruction* 14, no. 3 (2021): 717–36. https://doi.org/10.29333/iji.2021.14342a.

Mulyanto, Dedi, Muhammad Wahyudi, Arsyad Muhammad Ali Ridho, and Muhammad Zaki. "Utilization of Artificial Intelligence with Text-To-Speech Technology Based on Natural Language Processing to Enhance Arabic Listening Skills for Non-Native Speakers." *Alsinatuna* 10, no. 1 (2024). https://e-journal.uingusdur.ac.id/alsinatuna/article/view/7952.

P, Matan, and P. Velvizhy. "A Comprehensive Review of Expressive Text-To-Speech Systems and Its Advancements and Challenges." *2025 International Conference on Inventive Computation Technologies (ICICT)*, April 23, 2025, 49–54. https://doi.org/10.1109/ICICT64420.2025.11004848.

Reddy, V. Madhusudhana, T. Vaishnavi, and K. Pavan Kumar. "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning." *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, July 19, 2023, 657–66. https://doi.org/10.1109/ICECAA58104.2023.10212222.

Roussel, Stéphanie, André Tricot, and John Sweller. "The Advantages of Listening to Academic Content in a Second Language May Be Outweighed by Disadvantages: A Cognitive Load Theory Approach." *British Journal of Educational Psychology* 92, no. 2 (2022): 627–44. https://doi.org/10.1111/bjep.12468.

Sadoski, Mark, and Allan Paivio. *Imagery and Text: A Dual Coding Theory of Reading and Writing*. 0 ed. Routledge, 2012. https://doi.org/10.4324/9781410605276.

Saussure, Ferdinand De. *Writings In General Linguistics*. Oxford University PressOxford, 2006. https://doi.org/10.1093/oso/9780199261444.001.0001.

Sirmokadam, Sumukh. "Speech To Text for Data Entry—Opportunities and Challenges." In *Data Management, Analytics and Innovation*, vol. 137. Lecture Notes on Data Engineering and Communications Technologies. Springer Nature Singapore, 2023. https://doi.org/10.1007/978-981-19-2600-6_25.

Tilwani, Shouket Ahmad, Balachandran Vadivel, Yrene Cecilia Uribe-Hernández, Ismail Suardi Wekke, and Mir Mohammad Farooq Haidari. "The Impact of

Using TED Talks as a Learning Instrument on Enhancing Indonesian EFL Learners' Listening Skill." *Education Research International* 2022 (March 2022): 1–9. https://doi.org/10.1155/2022/8036363.

Tong, Shelley Xiuli, Kembell Lentejas, Qinli Deng, Ning An, and Yanmengna Cui. "How Prosodic Sensitivity Contributes to Reading Comprehension: A Meta-Analysis." *Educational Psychology Review* 35, no. 3 (2023): 78. https://doi.org/10.1007/s10648-023-09792-8.

Vogel, Adam P., Caroline Spencer, Katie Burke, et al. "Optimizing Communication in Ataxia: A Multifaceted Approach to Alternative and Augmentative Communication (AAC)." *The Cerebellum* 23, no. 5 (2024): 2142–51. https://doi.org/10.1007/s12311-024-01675-0.

Xiao, Yanling. "The Impact of AI-Driven Speech Recognition on EFL Listening Comprehension, Flow Experience, and Anxiety: A Randomized Controlled Trial." *Humanities and Social Sciences Communications* 12, no. 1 (2025): 425. https://doi.org/10.1057/s41599-025-04672-8.

Zhang, Yanhui, and Brian MacWhinney. "Using Diagnostic Feedback to Enhance the Development of Phonetic Knowledge of an L2: A CALL Design Based on the Unified Competition Model and the Implementation with the Pinyin Tutor." *Language Testing in Asia* 13, no. 1 (2023): 35. https://doi.org/10.1186/s40468-023-00232-6.

Zhong, Ruojun, and Yong Zhao. "Education Paradigm Shifts in the Age of AI: A Spatiotemporal Analysis of Learning." *ECNU Review of Education* 8, no. 2 (2025): 319–42. https://doi.org/10.1177/20965311251315204.

Zhu, Xinfa, Yi Lei, Tao Li, et al. "METTS: Multilingual Emotional Text-to-Speech by Cross-Speaker and Cross-Lingual Emotion Transfer." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32 (2024): 1506–18. https://doi.org/10.1109/TASLP.2024.3363444.

Biography of Authors

Asep Sunarko earned his Bachelor's degree in Arabic Language Education from UIN Saifuddin Zuhri Purwokerto, then pursued a Master's degree in Arabic Language Education at UIN Maulana Malik Ibrahim Malang, and a Doctoral degree in Islamic Studies with a concentration in Arabic Language Education at UIN Saifuddin Zuhri Purwokerto. His academic focus centers on the development of Arabic language instructional design and innovative Arabic teaching models in schools and Islamic boarding institutions (pesantren). He can be contacted at asepsunarko@unsiq.ac.id.

Muhamad Solehudin earned his Bachelor's degree in Islamic Law (HKI) from Darullughah Wadda'wah Islamic College, his Master's degree in Islamic Law (HKI) from UNSURI, and his Doctoral degree in Arabic Language Education from UIN Maulana Malik Ibrahim Malang. He can be contacted at muhamadsolehudin@uiidalwa.ac.id.

Nurin Sakinah is an alumna of Maktab Mahmud Yan, Kedah, Malaysia. She continued her studies at the International Islamic University Malaysia (IIUM), majoring in Islamic Revealed Knowledge Studies. n.sakinahnazli@live.iium.edu.my.