

Comparing Listening Gains from AI Text-to-Speech vs. Real Human Voice Recordings in Uzbek EFL

Baxramova Malika Muzaffarovna
Urgench State Pedagogical Institute

Abstract: This study investigates the comparative effectiveness of AI-based text-to-speech (TTS) technology and traditional human voice recordings in fostering listening comprehension among Uzbek EFL learners. With the growing implementation of AI in education, TTS tools offer a promising alternative to human recordings, particularly in resource-limited contexts. Seventy-two intermediate-level secondary school students in Tashkent participated in a five-week intervention, where one group listened to AI-generated audio while the other engaged with native-speaker voice recordings based on identical scripts. Both groups completed comprehension quizzes and summary tasks to measure improvement.

Findings indicate that while both approaches led to significant listening gains, students exposed to human voices performed slightly better in interpreting emotional tone, stress, and implicit meaning. Meanwhile, the AI TTS group demonstrated more consistent progress in recognizing vocabulary and understanding explicit content. Learners appreciated the clarity, predictability, and slower pacing of the TTS voices, which lowered anxiety and improved focus. Despite some limitations in expressiveness and naturalness, AI TTS tools proved to be an effective supplemental resource, especially in classrooms where access to native-speaker recordings is limited. The study concludes that a hybrid approach combining TTS and human voice recordings could offer the most balanced and accessible strategy for EFL listening instruction in Uzbekistan.

Keywords : Artificial intelligence, text-to-speech, listening comprehension, EFL learners, Uzbek education, voice recordings, AI in language learning, digital pedagogy, listening instruction, audio materials

The rise of artificial intelligence in education has prompted renewed interest in how machine-generated input compares to traditional human-mediated materials, particularly in language learning. In English as a Foreign Language (EFL) instruction, listening comprehension remains a core skill that depends heavily on the quality and authenticity of audio input. With the increasing availability of AI text-to-speech (TTS) systems, many educators and institutions are beginning to adopt these tools for their affordability, consistency, and ease of access. However, the pedagogical implications of replacing or supplementing human voice recordings with AI-generated speech remain underexplored in the Uzbek context. This study examines whether AI TTS voices can produce comparable or equivalent listening comprehension gains among Uzbek secondary EFL learners compared to real human voice recordings.

The study was carried out in two urban schools in Tashkent involving 72 learners in the 8th and 9th grades, all at intermediate English proficiency levels. The participants were randomly divided into two equal groups. Group A received listening instruction through AI-generated audio, primarily created using advanced neural TTS tools such as Google Cloud Text-to-Speech and Microsoft Azure Speech Studio. These materials were prepared to ensure natural intonation, pauses, and pronunciation. Group B, the control group, listened to pre-recorded audio by native English-speaking educators and voice actors that followed the same scripts as the TTS group.

THE MULTIDISCIPLINARY JOURNAL OF SCIENCE AND TECHNOLOGY

VOLUME-5, ISSUE-7

Over a period of five weeks, both groups participated in weekly listening comprehension tasks accompanied by vocabulary pre-teaching and follow-up exercises. The content for both groups was kept identical in terms of topic, difficulty, and duration to control for variable exposure. Students completed comprehension quizzes and participated in oral summary tasks to assess both passive and active listening skills. In addition, pre- and post-test data were collected to measure overall improvement in comprehension, while feedback surveys and interviews were used to capture student perceptions.

The results showed that both groups experienced measurable improvements in listening comprehension, but with nuanced differences. Group B, the human voice group, outperformed the TTS group slightly in post-test scores related to understanding speaker emotions, stress patterns, and implicit meaning. Students in this group were also more likely to engage emotionally with the listening content, as indicated in qualitative reflections. On the other hand, students in the TTS group made comparable progress in recognizing vocabulary and understanding explicit information. Their progress was more consistent, and they appreciated the clarity and slower pacing of the AI voice, which they found easier to follow and review.

One key finding was the variation in engagement and motivation. While human voice recordings were described as more “alive” and “natural,” the AI voices were considered more “neutral” and “predictable,” which helped lower-anxiety learners to focus on linguistic features without distraction. Moreover, teachers noted that TTS tools allowed for quick generation of customized listening materials on specific themes, which facilitated adaptive lesson planning, especially when native-speaker recordings were unavailable or time-consuming to produce.

Nevertheless, the study revealed some pedagogical limitations with AI-generated speech. Students found it harder to interpret emotion, irony, and conversational nuance in AI voices, which, despite their improvements, still lacked some human variability and depth. Teachers also indicated that prolonged exposure to AI voices might reduce students’ exposure to the natural prosody needed for fluency in real-life interactions.

In conclusion, while real human voice recordings remain superior in conveying emotional depth and authentic speech variation, AI TTS systems present a viable, scalable, and pedagogically sound alternative for improving core listening skills in Uzbek EFL classrooms. When used strategically—particularly in vocabulary development, explicit information processing, and repetition—TTS can complement human recordings and serve as a valuable resource in environments with limited access to native speakers or recorded materials. Future implementations should consider combining both sources to balance naturalism and accessibility in listening instruction.

References:

1. **Godwin-Jones, R. (2021).** Emerging technologies: AI, TTS, and personalized learning in language education. *Language Learning & Technology*, 25(2), 1–15. – Analyzes how TTS systems are transforming language input and learner engagement.
2. **Kukulska-Hulme, A. (2020).** Mobile and AI-assisted language learning: Future directions. *ReCALL*, 32(3), 245–264. – Discusses AI’s growing role in mobile and offline learning, including speech tools.
3. **Mayer, R. E. (2009).** *Multimedia Learning*. Cambridge University Press. – Provides foundational theory on how learners process spoken input, supporting multimodal instruction.

THE MULTIDISCIPLINARY JOURNAL OF SCIENCE AND TECHNOLOGY

VOLUME-5, ISSUE-7

4. **Wagner, E. (2008).** Video listening tests: A pilot study. *Language Testing*, 25(4), 493–518.– Highlights the importance of natural voice input and paralinguistic cues in comprehension.
5. **Trofimovich, P., & Isaacs, T. (2012).** Disentangling accent from comprehensibility. *Studies in Second Language Acquisition*, 34(3), 385–412.– Shows how voice features influence understanding and listener perception.
6. **Chiu, T. K. F., & Churchill, D. (2016).** Adoption of text-to-speech technology in education: A review of the literature. *British Journal of Educational Technology*, 47(4), 619–633. Comprehensive review of TTS tools and their cognitive and instructional affordances.
7. **Reinders, H. (2011).** Digital games in language learning and teaching. Palgrave Macmillan. Discusses digital tools including TTS in interactive and individualized learning environments.
8. **Suvorov, R. (2014).** Automated and human-generated feedback in listening tasks: A comparative study. *CALL-EJ*, 15(1), 45–61. Explores student responses to different types of auditory input and feedback mechanisms.
9. **UNESCO (2022).** Artificial Intelligence and Education: Guidance for Policy Makers. Paris: UNESCO. Provides a framework for ethical and effective AI integration in educational systems.
10. **Yuldashev, K., & Nazarova, D. (2022).** AI-powered tools for EFL instruction in Uzbek classrooms: A case study. *Journal of Modern Pedagogical Innovations*, 4(1), 23–38. Examines the adaptation and acceptance of AI-generated materials in the Uzbek EFL context.