YEARBOOK OF UNWE (2025), ACADEMIC PUBLISHING HOUSE UNWE
ISSN (print): 1312-5486; ISSN (online): 2534-8949; http://unwe-yearbook.org/en/
DOI: https://doi.org/10.37075/YB.2025.1.10

THE CONNECTION BETWEEN AUDIOBOOKS FOR LEARNING A FOREIGN LANGUAGE AND ARTIFICIAL INTELLIGENCE

Teodora Genova[1]
e-mail: t.genova@unibit.bg

Abstract

This paper aims to present the capabilities of artificial intelligence (AI) in the creation of audiobooks and how the production of audiobooks is changing with the help of new technologies in the face of AI. A short example of creating a sound file generated by AI using Text to Speech technology is given to illustrate the potential of AI in the field. On the other hand, in particular, the author of the paper also focuses on the capabilities of AI in the creation of audiobooks that are used in learning a foreign language compared to the audiobooks read by a human voice, and the ways in which foreign language learning is facilitated and personalized in this process.

Keywords: audiobook, artificial intelligence (AI), foreign languages, sound file, audio content

JEL: Z00

Introduction

The Introduction is a synthesized handout on the subject of the publication for the following sample questions: problem; existing solutions and degree of examination of the problem; limitations of the study; expected results of the author's study. Audiobooks, along with their initial function as a tool for easier access to books for blind people, are also used for learning foreign languages. Audiobooks influence the development of language competencies for listening, speaking, vocabulary enrichment, smooth reading, correct pronunciation and language fluency. Audiobooks are an example of audio-visual tools and are a part of mobile-assisted language learning, giving sensory perception to the information, thus providing a multisensory experience of the book's content (Genova, 2022).

Throughout the existence of audiobooks, the main question that has been at the forefront is whether listening to audiobooks can actually be perceived as reading, or as Rubery (2016) claims, "at the beginning of the creation of audiobooks, the main question was whether we read a book with our eyes or with our ears, and now in the modern technological world, the main question is whether the audio-

[1] Assoc. Prof., PhD, Department of Comprehensive Subjects, University of Library Studies and Information Technologies, Bulgaria, ORCID 0000-0002-7803-9461

book is read by a human or a robot, i.e. a voice generated by artificial intelligence (AI)" (Rubery, 2016).

On the website www.profit.bg we find that "audiobooks are also another market that AI can shake up. AI-generated voices are on the verge of taking the microphone away from the actors who read the works" (2023). From here occurs the question whether the listeners on audiobooks would prefer to listen to an audiobook, voiced by AI, not by professional narrators, so as the sound of the audiobook and the reading quality of the text are some of the main criteria when choosing an audiobook, especially when this is done for the purpose of learning a foreign language (Genova, 2022).

Several opinions stand out on this issue from parties involved in the research and production of audiobooks. On the one hand, many people still do not like listening to audiobooks voiced by AI, as they consider it too artificial and lack the feeling of creating a connection between the listener and the voice. Rubery points out that "most people find it difficult to listen to a non-human reading for more than a minute. Too often these voices lack the expressiveness of a real person. While human narrators have thought about the most appropriate way to read a story, mechanical narrators read every script in the same way" (Rubery, 2016).

On the other hand, according to Rubery (2016), for blind people, synthetic voices are preferable precisely for this reason. These listeners prefer computer-generated voices precisely because they do not perform the text; the artificial voice does not distract them from the content of the book, nor does it make them stare at the speaker's oddities; computer-generated voices do not interfere with the listener's own interpretation of the text.

With the development of new technologies and AI, the improvement of synthetic voices is expected to increase and therefore listeners will gradually adapt to them, thereby reducing their resistance. The future trend is to protect the rights of professional narrators in audiobooks by cloning their voices, thus ensuring and preserving the natural sound of the human voice.

The ways AI is changing the world of audiobooks

In our time, strongly dominated by new technologies, which are immensely influenced by the use of AI, the world of audiobooks does not remain unaffected. With the technological development of humanity, the connection between AI and audiobooks is becoming increasingly evident with AI serving to make the creation and use of audiobooks easier and cheaper, and facilitates their personalization. Examples of this modern trend are Google Play Books, as "the company offers publishers the ability to create books with an automated narrator, as long as they own the rights to the audiobook" (Profit.bg, 2023).

On the following lines several areas in which AI is influencing the change in audiobooks are outlined:

1. The main AI intervention that Rubery argues about is voice synthesis or the so-called Text -to-Speech (TTS) technology, using a robotic voice that is generated by AI. In this way, the written text of the audiobook is not read by professional narrators, as has been the case until now, but AI converts the text into natural-sounding speech. There are already a sufficient number of modern TTS systems (such as those from Google, Amazon, Microsoft, and others) that make audiobooks sound almost like real people with intonation, emotion, and pauses. This is due to "the text analysis module, which processes the input text and divides it into linguistic units such as phonemes, words, and sentences. This analysis is essential for determining the pronunciation, intonation, and stress of the generated speech… After text analysis, the next step in TTS is speech synthesis. This process involves generating the audio wave form that corresponds to the analyzed text. Machine learning models are trained on large datasets of text and corresponding speech recordings to learn the mapping between text inputs and audio outputs. By capturing the nuances of human speech, these models can produce high-quality synthetic voices that sound natural and expressive" (EITCA, 2024).

2. Thanks to TTS the creation of audiobooks is much easier and more accessible even to audiobook listeners themselves, which does not cost as much as listening to an audiobook that has been purchased or downloaded from the Internet and read by a human voice. In addition, the audiobook listener can personalize their audio file creation by choosing a male or female voice, pace, intonation and reading style. In this regard, AI takes over the main function of the narrator in audiobooks, which can be of great benefit to independent authors with limited budgets and the rapid translation and dubbing of books into different languages without the presence of a studio and narrator. This is especially useful when learning a foreign language, as different features of the audiobook can be selected according to different levels of proficiency in the foreign language, i.e. a slow female voice can be selected for beginner learners and a faster and more conversational male voice for advanced learners.

3. Another major role that AI plays in the development of audiobooks is automatic editing and the review of text, with AI having the ability to analyze a text and automatically correct errors, suggest a smoother structure for a better audio experience, and create automatic summaries or annotations.

4. Fourth, AI helps to translate books into different languages faster and easier and immediately narrates them in another language, which in turn leads to the global distribution of book content. This facilitates access to audiobooks in many languages. Since AI can read audiobooks in different languages, this enriches the thematic and genre spectrum when finding a variety of audiobook content from classic fairy tales, novels, to educational texts.

5. The latest direction in which audiobooks are developing with the help of AI is the creation of the so-called interactive audiobooks, in which the listener chooses how the story develops – this is something like an audio version of "choose your own adventure". In interactive audiobooks, the listener can "participate" by making decisions, asking questions, or influencing the development of the story, and AI reacts in real time. The main tasks performed by AI here are generating new storylines in real time; recognizing the emotions, interests and preferences of the listener; responding to voice commands and creating conversations with the characters as if the listener were actually participating in the story. Examples of interactive audiobooks created using AI include FableLabs – a platform with interactive stories similar to games; Novel Effect – synchronizes animations, sounds and reactions with the narrator's voice; Choose Your Adventure (AI-like versions) – the user chooses what to happen; Replika Stories – conversational stories with AI characters who "talk" to you, etc. Interactive audiobooks are suitable for children when learning through play and developing imagination; people who are learning a foreign language to practice the language through interaction; readers who like dynamic, unconventional stories; people with disabilities thanks to the voice control provided by AI.

Example of creating a sound file using AI

In this part, the steps involved in creating an AI-generated audio file of a short text similar to the text of an audiobook are briefly discussed. The purpose of this part of the paper is to illustrate in practice the enormous potential of AI in voicing audiobooks in different languages, using English here. For this purpose, one of the numerous AI audio content creation platforms ElevenLabs is used. This online platform, which uses AI through Text to Speech technology to create audiobooks, offers one of the most realistic-sounding robotic voices currently in the field.

The first step is for the user to register with the platform online and create a free account. Then, the user must go to the Text to the Speech or Speech Synthesis section, where they can select a voice. The ElevenLabs platform offers a variety of natural-sounding voices that can be selected from a rich library with a variety of sounds, e.g. dark, scary, low, romantic, calm, etc. From there, a voice is selected that matches the desired style – for example, gentle and calm.

The second step involves entering the text into the text box. For better sounding, the text can be divided into paragraphs or scenes.

The third step is related to the settings of the robotic voice, and there are numerous options. In addition to the type of voice, the platform can adjust the voice with the following functionalities – model, speed, stability, similarity and style exaggeration, emotional intonation, which give opportunity for active choice at preferences for listening on audio content.

The last step is related to generating an audio file and once the audio content generation process is complete, the finished MP3 file can be downloaded.

AI-generated audiobooks and language learning

As has been made clear earlier in this report, AI is significantly changing the way we create and use audiobooks, and this has a direct impact on the way we learn foreign languages. Here, several guidelines can be drawn that emerge when using audiobooks created by AI compared to audiobooks read by a human voice in learning a foreign language, as presented in Table 1. The guidelines considered are: pronunciation and intonation; emotional engagement; accessibility and personalization; example of natural-sounding language; price and accessibility.

Table 1: Comparison between audiobooks read by a human voice and audiobooks created by AI

1. Pronunciation and intonation	
Human voice	AI voice
It usually offers natural pronunciation with rich intonation, rhythm, and melody of speech, which helps with better acquisition of the foreign language.	Despite advances in AI, new AI systems often lack subtle emotional and intonation depth. Some nuances in stress and rhythm still sound artificial.
2. Emotional commitment	
Human voice	AI voice
The human voice can convey emotion, emphasis, and drama, which improves the understanding of context and makes it easier to remember.	AI voices are getting better at simulating emotions, but they often sound more mechanical or flat, especially in complex dialogue or narration.
3. Accessibility and personalization	
Human voice	AI voice
The human voice can read a limited number of titles, as recordings take time and resources.	AI enables the mass generation of audiobooks, including in less common languages or levels. You can even customize the voice or pace.
4. Example of natural-sounding language	
Human voice	AI voice
It provides a realistic role model – what the language sounds like, the way it is spoken by its native speaker.	If an AI model is well trained, it can be useful for phonetic learning, but it is not always completely reliable for imitation.
5. Price and availability	
Human voice	AI voice
Often more expensive and more limited in volume.	Much more accessible and widespread. Some platforms offer automatic AI reading of every e-book.

Source: Composed by the author

The comparison between audiobooks read by a human voice and those read by AI shows that the first type of audiobooks, in which listeners hear a natural human voice, are more suitable for beginner learners, since at this stage a model for linguistic imitation of natural-sounding speech is sought and the learners of a certain foreign language themselves have not yet become accustomed to and mastered the prosodic characteristics of the language such as rhythm, intonation, stress, etc. In addition to this advantage, audiobooks read by professional narrators also carry the characteristic of emotional engagement, with which they still attract their listeners, who feel more comfortable and trusting, listening to a human voice, which helps them to feel the content of the text in greater fullness and depth and provides a more comprehensive understanding of the context of the story. Last but not least, an advantage of this type of audiobook is the fact that they can be used to subtly improve pronunciation by learners at higher levels of language proficiency, where the subtle nuances of the language can be felt.

On the other hand, with the improvement and expansion of AI, audiobooks read by a robotic voice also have their advantages and can be used to support language learners, as they are a more accessible alternative to audiobooks in which a human voice sounds, in cases where the listener needs to build on their knowledge of a foreign language and be exposed to the constant influence of foreign language speech. Audiobooks generated by AI can support learners in developing their listening and reading comprehension skills on a larger scale due to the mass generation of audiobooks combined with the function of selecting and personalizing audio content. Due to their accessibility, widespread distribution and budget-friendly nature, audiobooks read by a robotic voice can be reliable allies of language learners at every stage of the development of their language skills, as they offer automatic reading of each e-book and provide access to a wide variety of languages.

Along with the advantages of both types of audiobooks, there are also a number of disadvantages. Audiobooks read by AI need further improvement in order to achieve smoother and more natural-sounding speech and subtle emotional and intonation depth. As for audiobooks read by professional narrators, one can only speculate whether they will be the preferred way to listen to audio content in the future, despite their limitations in terms of creating a variety of styles and genres, volume, and their price on the market.

Personally, the best strategy for learning a foreign language with the help of audiobooks is to combine both types – using the human voice for basic or important texts, and the robotic voice for more routine listening.

There are already enough useful resources and platforms for audiobooks, read by both human and AI voices, specially selected for learning a foreign language. The first type of audiobooks includes Audible (from Amazon), Librivox, Book2

from Goethe Verlag, etc., where you can find free audiobooks in different languages in different genres and separate sections with books for learning foreign languages with lessons. Examples of the second type of audiobooks include Google Play Books, Speechify, AI Text to Speech Natural Reader, etc., which offer automatic reading by AI voice, an AI application for converting text to audio format, support for a large selection of languages, etc.

An empirical example of the effectiveness of an AI-based application, which incorporates audiobooks for learning English as a foreign language, is ELSA Speak. It "utilizes artificial intelligence (AI) to help users improve their pronunciation skills". In this study, sound evidence is given for the improvement of EFL students' pronunciation skills that use ELSA Speak (experimental group) compared to those using U-Dictionary (control group). "This study demonstrates that the ELSA Speak application significantly improves EFL students' pronunciation skills compared to traditional methods" (Permatasari & Lubis, 2024, p. 210).

Apart from this study, there are also some other sources which deal with the potential of AI-generated resources for learning a foreign language, for example Baxramova (2025), Woo & Choi (2021), etc., but their number is still scarce and not-on-the-topic, because the theme of how audiobooks generated by AI enhance students' language acquisition remains quite nascent to this day and there is plenty of room for future research including empirical studies on evaluating AI platforms directly integrating audiobooks for learning a foreign language offering automated feedback, personalization, or adaptive features.

Conclusion

This paper examines the connection between audiobooks for learning a foreign language and AI and the development of this type of audio format in the form of audiobooks read by professional narrators and those generated by AI. The author of the paper attempts to examine the attitude towards both types of audiobooks, proposing a short consideration of the development of audiobooks in our time, which has been strongly influenced by the use of AI in the creation of audiobooks. An example of creating an audio text and reading it by AI using the Text to Speech technology is proposed, which fundamentally changes the development of the audiobook market, bringing countless benefits to listeners of audio content such as personalization of the audio experience, automation of audiobook production, ease, accessibility and expansion of the scale of use of audiobooks in different languages. Thanks to AI, a new type of audiobook is emerging – the so-called interactive audiobooks, which provide a radically different sensory experience of the text, with the listeners themselves becoming part of and actively participating in the plot of the story.

It is an indisputable fact that audiobooks are used for learning foreign languages, and with the development and improvement of AI, more and more listeners would trust audiobooks read by AI on their path to mastering and improving their foreign language skills due to the growing number of advantages of audiobooks generated by AI, reinforcing the trend towards digitalization of the media and changing the way we can enjoy a literary text in a foreign language.

Notes

In this report, the author has used answers to questions asked to the AI platform Chat GPT.

References

Baxramova Malika Muzaffarovna. (2025). Effectiveness of AI-generated Listening Materials vs. Textbook Audio in EFL Learning. Ta'lim Innovatsiyasi Va Integratsiyasi, Volume 49(1), pp. 53-55

Генова, Т. (2022). Ролята на аудиокнигите като медия в чуждоезиковото обучение по английски език, Монография, София: Академично издателство „За буквите – О писменехь". (Genova, T., 2022, Rolyata na audioknigite kato media v chuzhdoezikovoto obuchenie po angliyski ezik, Monografia, Sofia: Akademichno izdatelstvo „Za bukvite – O pismenehy").

EITCA. (2024). Какво е текст към реч (TTS) и как работи с AI?, [online], (EITCA, 2024, Kakvo e tekst kam rech (TTS) i kak raboti s AI), available at: https://bg.eitca.org/artificial-intelligence/eitc-ai-gcml-google-cloud-machine-learning/introduction/what-is-machine-learning/what-is-text-to-speech-tts-and-how-it-works-with-ai/

Profit.bg (2023). AI клонинги на човешки гласове идват за дял от пазара на аудиокнигите, [online], (Profit.bg, 2023, AI kloningi na choveshki glasove idvat za dyal ot pazara na audioknigite), available at: https://profit.bg/article/2024012516122395042

Permatasari, S., & Lubis, Y. (2024). Enhancing Pronunciation Skills in EFL Students through the ELSA Speak Application, Indonesian EFL Journal, Volume 10(2), pp. 205-212.

Rubery, M. (2016). The Untold Story of the Talking Book, Harvard University Press, p. 380.

Woo, J. H. & Choi, H. (2021). Systematic Review for AI-based Language Learning Tools, Journal of Digital Contents Society, Volume 22(11), pp. 1783-1792.