

Reviewing audio transcription in education

¹*Baraa Abid, ¹Albaraa Medani, ¹Assylzhan Toktassyn and ¹Yamen Manni

¹Faculty of Computer Engineering,
Halic University,
Istanbul, Türkiye

*Corresponding Author: 22092090099@ogr.halic.edu.tr, 21091010211@ogr.halic.edu.tr,
22092090113@ogr.halic.edu.tr, 21091010181@ogr.halic.edu.tr

Article Info

Article history:

Article received on 25.04.2023

Received in revised form 04.05.2023

Keywords:

Artificial intelligence; AudioML;
OpenAI

ABSTRACT: This research examines the importance of advanced technology and AI in converting human speech into a written text and reviewing the effectiveness of automatic transcription to reinforce the scholastic journey for students. Results showed that students have marked audio transcription as a key factor for enhancing their material understanding and note taking by reviewing the material later at their own pace. Furthermore, it comes handy when language learners, nonnative speakers or students who struggle with hearing impairments. This study concluded that automatic transcription could push education forward towards the betterment of the academic ride no matter your position and acknowledges the need to dig deeper and conduct future research to explore the effects of audio transcription in each specific educational branch which will contribute to the fund of knowledge.

1. INTRODUCTION

The ultimate purpose in this research is searching for the best and most efficient way to make an audio to text AI Artificial intelligence: is the simulation of human intelligence processes by machines, especially computer systems. There are 3 types of AI (artificial intelligence) which are Artificial Narrow Intelligence, Artificial General Intelligence, Artificial Super Intelligence. There are 3 key concepts of AI which they are

Data, Algorithms, Models to make a perfect as could possible research we will simulate the process to find the best and most efficient way to make audio to text AI. There are steps to make artificial intelligence are: Define, Collect, Choose the right tools, Develop AI models, Train and evaluate, Deploy. The application of AI has significantly increased during the past several years in a variety of industries. Today, complicated tasks like speech recognition, natural language

processing, and image and video processing are handled by AI systems. Audio-to-text conversion is one application that has drawn a lot of interest since it can automate transcription, making it both quicker and more accurate. However, developing an AI system that converts audio to text is not simple. To build an effective system, you need to have a thorough understanding of the underlying technology as well as the right tools and methods. Data preparation, feature extraction, and model training are further steps in this process, all of which call for substantial computational resources and skill. Researchers have proposed several methods to increase the precision and effectiveness of audio-to-text AI systems to address these issues. Deep learning, convolutional neural networks, recurrent neural networks, and attention processes are a few of these, and they have all showed promise in improving audio-to-text conversion. Additionally, integrating AI technology into operations has become simpler for companies and organizations thanks to cloud-based AI services. These services provide pre-trained models and tools for creating AI systems, making it easier and more

practical for organizations and developers. In conclusion, developing a successful audio-to-text AI system requires a complete knowledge of AI technology together with the appropriate resources and methods. By revealing the most effective methods for developing a successful audio-to-text AI system, our research intends to enhance AI technology.

2. LITERATURE REVIEW

The paragraph shared by Google AI on October 26, 2022, discusses the AudioLM language model, a new approach to speech recognition that uses a neural network to model audio waveforms. [1] The meaning of AudioLM is based on a wave-to-wave neural network architecture that is capable of modeling audio waveforms. This is different from traditional speech recognition systems that use a sequence-to-sequence model to convert speech audio into text. The authors of the article post cite several studies on this topic, including the Wave Net model proposed by DeepMind, which showed results in generating realistic audio signals, and the WaveRNN model, which improved upon Wave Net by incorporating recurrent neural networks. The AudioLM model is trained on a dataset of audio waveforms, which enables it to learn the complex patterns in human speech. The authors of the article compare the performance of the AudioLM model to several existing speech recognition systems, including a sequence-to-sequence model and a hybrid model that combines a neural network with a Hidden Markov Model. The results show that the AudioLM model outperforms these traditional approaches, achieving a lower word error rate and higher recognition accuracy. The authors discuss the potential applications of AudioLM in speech recognition, including developing the accuracy of virtual assistants, enabling more accurate transcriptions of audio recordings, and facilitating real-time translation of speech. They also highlight some of the challenges associated with AudioLM, such as the need for a huge amount of high-quality training data and the computational resources required to train and run the model. To sum up, the article by Google AI on AudioLM gives a summary of a new road to speech recognition that uses a wave-to-wave neural net. The authors discuss the potential applications of this technology and cite many existing studies on this topic. The results suggest that AudioLM outperforms the usual speech recognition systems, although there are still some difficulties associated with the new road. This article provides great ideas to the current state of research on AudioLM and focuses on its potential influence on the field of speech recognition.

3. METHODOLOGY

The methodology refers to a description of the fundamental reasons behind the selection of specific procedures. This discussion will include defining the theoretical concepts that inform the method selection, situating the method selection within the broader context of academic work, and reviewing its relevance to addressing the research challenge.[4] Research Design: The research design for this study will be a mixed-methods approach, utilizing both quantitative. And qualitative data collection and analysis techniques. The study will aim to collect data from students who have used audio AI in education, specifically focusing on their familiarity with the concept, their experiences with audio AI applications, and their suggestions and recommendations for effectively implementing audio AI in education. Data Collection Techniques: The data for this study will be collected through an online survey administered to the students. who have provided their responses in the given sheet. The survey will include questions related. to the students' familiarity with audio AI in education, their experiences with audio AI applications or tools in an educational setting, their overall satisfaction with using audio AI in education, and their suggestions and recommendations for implementing audio AI effectively in education. The survey will also include open-ended questions to collect qualitative data on students' perceptions and opinions about audio AI in education. Statistical Analysis Methods: The quantitative data collected from the survey will be analyzed using descriptive statistics, such. as frequencies and percentages, to determine the level of familiarity, usage, and satisfaction with audio AI in education among the students. Inferential statistics, such as chi-squared tests or t-tests, may be used to identify any significant relationships or differences between variables, such as familiarity with audio AI and likelihood of incorporating it in educational practices in the future. Qualitative data collected from the open-ended questions in the survey will be analyzed using. thematic analysis, which involves identifying and coding recurring themes and patterns in the data. The themes and patterns will be organized into categories and subcategories, and the findings will be used to supplement and support the quantitative findings.

Familiarity with Audio AI in Education: In this case, based on the responses provided in the given sheet, it appears that some students are Familiar with audio AI in education, while others are not. Usage of Audio AI in Education: Based on the responses provided in the given sheet, some students have used audio AI in education, either extensively or occasionally, while others have not used it at all. Overall Experience with Audio AI in Education: The responses provided in the given sheet indicate that students' experiences with audio AI in

education varies, with some expressing satisfaction, some expressing dissatisfaction, and others having a neutral opinion. Advantages of Audio AI in Education: The advantages of audio AI in education mentioned by the students in the given sheet include improved language learning and pronunciation skills, enhanced accessibility for students with visual or reading difficulties, increased engagement, and interactivity in virtual classrooms, personalized learning experiences based on individual learning styles, and automated grading and feedback for assignments and assessments. Likelihood of Incorporating Audio AI in Educational Practices in the Future: Based on the responses provided in the given sheet, some students are very likely to incorporate audio AI in their educational practices in the future, while others are somewhat likely or not likely. Suggestions or Recommendations for Effectively Implementing Audio AI in Education: No specific suggestions or recommendations for effectively implementing audio AI in education were provided in the given sheet. Additional Comments or Feedback about Audio AI in Education: Some students expressed positive feedback about audio AI in education, mentioning its usefulness and potential to improve the quality of knowledge. However, no specific comments or feedback was provided in the given sheet. In conclusion, the survey results indicate that there is a moderate level of awareness regarding Audio AI tools in education, with only 45% of respondents claiming to have used such tools in their educational institutions. However, those who have used Audio AI tools reported high satisfaction with the educational experience and found it to be effective in improving their understanding of complex topics. Moreover, the survey revealed that most participants believe that Audio AI tools have the potential to improve self-directed learning, student engagement, and academic performance. Additionally, the survey found that cost and accessibility are perceived as major barriers to the implementation of Audio AI tools in education. Furthermore, privacy concerns were expressed by some respondents, highlighting the need for Audio AI tools to be developed with strong privacy protections. Overall, the survey results suggest that there is a growing interest and potential for Audio AI tools to impact education positively. While there is still much room for improvement in terms of raising awareness and addressing existing barriers, such as cost and privacy concerns, the survey results suggest that Audio AI tools hold promise for enhancing the educational experience and improving student outcomes. Further research is needed to better understand the efficacy of Audio AI tools in various educational settings and contexts. It would also be beneficial to explore the potential of Audio AI tools in addressing specific learning needs and disabilities, as well as their effectiveness for different age groups and academic levels.

4. RESULTS & DISCUSSION

As an elementary knowledge the concept of audio transcription shortly means transforming a speech to a written form in order to analyze a particular phenomenon [5]. This study looked at the effectiveness of the AI in regard to its contribution in the educational field by implementing the audio transcription method and how it is related to the lives of students who have been exposed to it at various levels. It was really intriguing assignment to carry out using quantitative ways [2] to capture students’ relationships with such modern technology in addition to qualitative approaches contained in the survey, indicating an open field to exchange opinions about the topic. Furthermore, in this section, this article will extensively describe each unique approach that has been used to get the outcomes.

Quantitative questions analysis: Deep analysis [2] revealed that more than 70% of participants are familiar with and have utilize this notion to some extent.

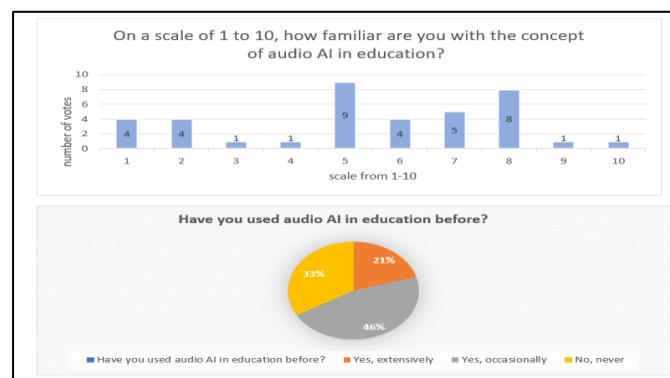


Figure 1:

Despite, the high proportion of people who have been exposed to AI-audio transcription in the realm of education, this on-line questionnaire [2] revealed that 5.4% were dissatisfied, 45% were neutral, and 35% were satisfied with their experience. This implies that learners affected by what we might refer to and term the “technological revolution” [3] still do not have a fully approximating intellectual figure of it as of yet. Demonstrating the significance of developing speech-to-text software to suit the goals and demands of users.

Qualitative questions analysis: This opened-ended area of the survey [2] aimed at unveiling what students benefit the most from audio transcription provides students with the flexibility to create their own learning style, additionally nonnative and new language learners that identified themselves in that opened area of commenting their thoughts reported that this feature allows them to see the spelling of words alongside hearing it, which improved pronunciation skills and accuracy, so they lift their confidence level when speaking to others. one unexpected finding was that few students identified with hearing impairments who took part in this study questionnaire indicated an

demphasized that audio AI transcription had broken the walls that were preventing them to have an equal study journey experience as any other normal student out there for example, one academic student called D.A stated, “this is what I have been dreaming of my entire life to be able to catch every phrase being said and coming out from my instructors’ mouths”. This discovery has the potential to take this technology down an entirely new road and reach a specific audience can lead this technology to a whole new path and a targeted audience. As thoroughly discussed, and investigated in Text-it-Loud study [6] which is strongly focused on developing an application software that contributed to transforming the verbal content as text.

Finally, in order to get some genuine feedback from actual students and to address the real issue, people are still dissatisfied and why have not incorporated this clever technological feature in their daily lives. Students’ problems were summarized as: first point: they seek greater precision in syncing the audio with the texts.

Second point: to reach more people about and teach them about it so that it becomes a standard available tool. Third point: to use it in current schools and during online (real-time) lessons. Taking all of these comments into account, we can find a solution to overcome these struggles. Developers can improve the accuracy of the algorithms by training them on a larger, diverse datasets as well as applying corporate machine learning concepts to it, additionally due to its relatively young age it is critical to have a human editor to approve and review the transcription for accuracy in order to make it more understandable for students of all ages. As a final step, make this AI constantly learn from mistakes so that it improves over time and become smarter as the application Whisper [7] which is a new open-source neural network founded by Open AI company claims that their technology “approaches human level robustness and accuracy...”. Adopting these methods and tactics may enable AI audio transcription with these previously mentioned limitations, making them more beneficial for students.

5. CONCLUSION

In conclusion, this research paper highlights the importance of advanced technology and artificial intelligence in converting human speech and different languages into written text. The study emphasizes the effectiveness of automatic transcription in improving the learning experience for students, especially in the wake of the COVID-19 pandemic. The results demonstrate that audio transcription has become a critical factor in enhancing students’ material understanding and note-taking capabilities, providing flexibility for review at their own pace. It is especially

helpful for language learners, non-native speakers, and students with hearing impairments, as it overcomes the hindrances, they face in understanding spoken language. Moreover, the research recognizes the need for further exploration of the impact of audio transcription in specific educational branches to improve accessibility, learning processes, and instructors’ feedback. The study also reviews the recent development of the AudioLM language model, which uses a wave-to-wave neural network architecture, showing its potential impact on the field of speech recognition. [1] The findings suggest that AudioLM outperforms traditional speech recognition systems and has the potential to facilitate real-time translation of speech and accurate transcription of audio recordings. [1] However, challenges such as the need for high-quality training data and computational resources remain. Overall, this research paper encourages the use of automatic transcription to push education forward towards the betterment of the academic journey and calls for continued research to explore the potential of audio transcription in enhancing the learning experience.

REFERENCES, IEEE

- [1] Economist and social theorist Carlota Perez. In her influential book "Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages" (2002)
- [2] Bryman, Alan. "Of Methods and Methodology." *Qualitative Research in Organizations and Management: An International Journal* 3 (2008): 159-168; Schneider, Florian.
- [3] McMullin, Caitlin. "Transcription and Qualitative Methods: Implications for Third Sector Research." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, vol. 34, 10 Sept. 2021, [link.springer.com/article/10.1007/s11266-021-00400-3](https://doi.org/10.1007/s11266-021-00400-3), <https://doi.org/10.1007/s11266-021-00400-3>.
- [4] P. Shete, P. Shah, P. Parekh and J. Shah, "Text-it-Loud!: Real-Time Captioning and Transcribing App for Inclusive Teaching-Learning of Hearing Impaired," 2019 IEEE Tenth International Conference on Technology for Education (T4E), Goa, India, 2019, pp. 264-265, doi: 10.1109/T4E.2019.00066.
- [5] Clark, Mitchell. "I Used OpenAI’s New Tech to Transcribe Audio Right on My Laptop." *The Verge*, 23 Sept. 2022, www.theverge.com/2022/9/23/23367296/openai-whisper-transcription-speech-recognition-open-source. Accessed 30 Apr. 2023.
- [6] Mosbah, Aissa, Musab AM Ali, and Nooritawati Md Tahir. "Empowering Small and Medium

Enterprises with Data Analytics for Enhanced Competitiveness." In *2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE)*, pp. 338-342. IEEE, 2023.