

## Audio Summarization of Podcasts

Devika Rani Roy<sup>1</sup>, Sakshi Narendra Patil<sup>2,\*</sup>, Ritika Baban Parate<sup>3</sup>, Tanaya More<sup>4</sup>

### Abstract

*Podcasts have emerged as a significant medium for disseminating information, sharing stories, and providing entertainment. As their popularity continues to soar, the sheer volume of available content poses a challenge for listeners seeking to efficiently consume information. In this context, creating and deploying an audio summarizer for podcasts becomes highly significant. This research paper delves into the motivation for creating such a tool, emphasizing the increasing need for concise and informative summaries that can help users navigate the vast array of podcast episodes more effectively. By providing an overview of the methodology used for summarization, the paper outlines the steps involved in distilling key points from lengthy audio content. It also delves into the technological foundations enabling audio summarization, such as advancements in natural language processing, machine learning algorithms, and speech-to-text technology. Additionally, the paper discusses potential applications of an audio summarizer, highlighting its utility for various user groups such as busy professionals, students, and casual listeners who aim to stay informed without investing extensive time. Overall, the development of an audio summarizer for podcasts represents a crucial innovation in enhancing the accessibility and efficiency of information consumption in the digital age.*

**Keywords:** Audio summarizer, podcast summarizer, natural language processing (NLP), automatic speech recognition (ASR), text-to-speech (TTS), content discovery, podcast transcription, text summarization, podcasting technology, audio processing

### INTRODUCTION

Podcasts have emerged as popular and influential media for delivering information, entertainment, and storytelling. However, the rapid growth of podcast ecosystems has given rise to several challenges for both podcast consumers and creators. The central problem addressed by an “Audio Summarizer for Podcasts” is the overwhelming abundance of podcast content and the associated difficulties in efficiently navigating this vast array of information. With thousands of new episodes released daily across diverse topics, listeners often find it daunting to identify and consume content that aligns with their interests and time constraints.

#### \*Author for Correspondence

Sakshi Narendra Patil  
E-mail: sakshipatil@kccemsr.edu.in

<sup>1</sup>Professor, Department of Information Technology, K.C. College of Engineering and Management studies and Research, Thane, Maharashtra, India

<sup>2-4</sup>Student, Department of Information Technology, K.C. College of Engineering and Management studies and Research, Thane, Maharashtra, India

Received Date: June 28, 2024  
Accepted Date: August 02, 2024  
Published Date: September 14, 2024

**Citation:** Devika Rani Roy, Sakshi Narendra Patil, Ritika Baban Parate, Tanaya More. Audio Summarization of Podcasts. Journal of Software Engineering Tools & Technology Trends. 2024; 11(3): 18–26p.

An audio summarizer for podcasts aims to alleviate these challenges by providing concise and informative summaries of podcast episodes. This tool leverages advanced technologies, such as natural language processing (NLP), machine learning, and audio synthesis, to extract and condense key information from lengthy audio files. The summarizer can identify the main themes, highlight important discussions, and present them in a shortened format, allowing listeners to quickly grasp the essence of an episode without committing to its full length.

---

For podcast creators, summarizers offer the benefit of increased discoverability and engagement. Creators can attract potential listeners who might be deterred by the duration of episodes but are interested in the content. Additionally, these summaries can be used for promotional purposes, providing a snapshot of an episode's value proposition.

Essentially, the audio summarizer represents a transformative solution for the podcasting landscape. It addresses modern listeners' need for efficient content consumption while supporting creators in reaching a broader audience. By bridging the gap between the extensive volume of podcast content and the limited time available to consumers, this innovative tool has the potential to revolutionize how we discover, engage with, and benefit from the vast pool of audio-based knowledge and entertainment available today.

## LITERATURE REVIEW

The literature review provides a comprehensive review of research and development related to the topic of this study, as shown in Table 1. This section discusses the background, key findings, methods, and conclusions of previous studies. This research aims to identify gaps in existing knowledge and provide a foundation for current research by analyzing various sources. The literature review was organized by topic to highlight various theories and practices explored by other researchers.

## PROPOSED SYSTEM

Creating an audio summarizer for podcasts involves a multistep methodology that includes transcription, NLP, summarization, and audio synthesis, as shown in Figure 1. Here is a detailed methodology [1].

### Data Collection

- *Podcast audio*: Collect the podcast episodes you want to summarize. These can be obtained from podcast host platforms or through web scraping.

### Audio Transcription

- *Automatic speech recognition (ASR)*: ASR technology is used to convert podcast audio into written transcripts. Popular ASR engines include Google ASR, Amazon Transcribe, and custom ASR.

### Preprocessing

- *Text cleaning*: The transcribed text is cleaned and preprocessed to remove punctuation, unwanted formatting, and noise.

### Natural Language Processing

- *Tokenization*: Tokenize the text into sentences and words for further analysis.
- *Named entity recognition (NER)*: Identify and extract named entities (e.g., people, places, organizations) from the text.
- *Part-of-speech tagging*: Annotate words with their grammatical categories (e.g., noun, verb, adjective).
- *Stop word removal*: Remove common stop words (e.g., "and," "the") to focus on meaningful content.
- *Text summarization*: Text summarization techniques are applied to generate a concise summary from the transcript.
- There are two primary types of text summarization. Extractive summarization involves selecting important sentences directly from the transcript. Abstractive summarization generates summaries by paraphrasing and rephrasing key contents.

**Table 1.** Literature review.

S.N.	Title of paper/website	Advantages	Limitations	Green IT
1.	<i>Title:</i> Audio summarization in real time for podcast, speeches and audiobooks.  <i>Authors:</i> Liya B.S., Seenuvasan V, Sathya Prakash K, Siva B [2]	Real-time summarization, applicable to various audio formats.	Limited dataset evaluation, lack of established efficiency benchmark.	Potential energy reduction by grasping key points quickly.
2.	<i>Title:</i> A survey on abstractive summarization techniques <i>Author:</i> Rachabathuni PK. [3]	Attempts abstractive summarization for more concise summaries.	Abstractive summarization can be challenging for audio.	A potential reduction in listening time saves energy, but model training and computation might negate it.
3.	<i>Title:</i> Extractive summarization of call transcripts. <i>Authors:</i> Biswas PK, Iakubovich A [4]	Leverages topic modeling and sentiment analysis for a comprehensive summary.	Extractive summarization may not capture conversational nuances.	Reduces listening time, potentially saving energy.
4.	<i>Title:</i> Podcast Summarization System <i>Authors:</i> Vanjari N, Pinjari S, et al. [5]	Utilizes neural networks for potentially higher accuracy summaries.	Significant computational resources are required, limited by training data.	High computational costs may negate potential energy savings.
5.	<i>Title:</i> SummarizeAI- summarization of the podcasts <i>Authors:</i> Khanna D, Bhushan R, et al. [6]	Integrates speaker diarization for identifying individual speakers, potentially improving summary clarity.	Speaker diarization can be challenging in noisy environments.	Reduces listening time, potentially saving energy.
6.	<i>Title:</i> Fundamental building block for all-optical scalable quantum networks. <i>Authors:</i> Lee SW, Ralph TC, Jeong H [7]	Explores abstractive summarization for more concise and informative summaries.	Abstractive summarization can be more prone to factual errors.	Reduces listening time, potentially saving energy.
7.	<i>Title:</i> PodSumm--Podcast Audio Summarization <i>Authors:</i> Vartakavi A, Garg A [8]	Focuses on identifying key topics for a structured summary.	May not capture all conversational aspects not directly related to the main topics.	Reduces listening time, potentially saving energy.
8.	<i>Title:</i> Dual attention model for citation recommendation with analyses on explainability of attention mechanisms and qualitative experiments <i>Authors:</i> Zhang Y, Ma Q. [9]	Combines attention mechanism and topic modeling for potentially more relevant summaries.	Complexity can increase computational cost.	The trade-off between potential energy savings and increased computational resources.
9.	<i>Title:</i> Video Transcript Extraction and Summarization Using Transfer Learning <i>Authors:</i> Mehta V, Deshpande T, et al. [10]	Leverages existing subtitles (if available) for efficient summarization.	Rely on subtitle quality and availability.	Reduces reliance on speech recognition, potentially saving energy.
10.	<i>Title:</i> Newspod: Automatic and interactive news podcasts <i>Authors:</i> Laban P, Ye E, Korlakunta S, et al. [11]	Segments the podcast into meaningful chunks for a structured summary.	Segmentation accuracy can affect summary quality.	Reduces listening time, potentially saving.

---

## Summarization Techniques

### Extractive Summarization

- *TF-IDF*: Calculate the importance of each sentence based on term frequency-inverse document frequency.
- *Text rank*: Apply graph-based algorithms to rank sentences based on their relationships within the text.
- *Bidirectional Encoder Representations from Transformers (BERT)-based models*: Utilize pre-trained language models such as BERT to score and select sentences.
- *Abstractive summarization: GPT-3 or Similar Models*: Use large language models such as GPT-3 to generate abstractive summaries.
- *Sequence-to-sequence models*: Training sequence-to-sequence models. (e.g., using transformers) for abstractive summarization.

### Audio Synthesis

- *Text-to-speech (TTS)*: Convert the generated text summary into audio using a TTS engine. TTS services such as Google Text-to-Speech or Amazon Polly can be used.

### Audio Post-processing

The audio editing trim adjusted and enhanced the synthesized audio to ensure a smooth and coherent listening experience. Add Introduction and Conclusions: Include brief audio introductions and conclusions to the summary for context and clarity.

### Evaluation

- *Evaluation metrics*: Assess the quality of the generated summaries using metrics such as the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) for text-based summaries or subjective user feedback for audio summaries.

### User Feedback and Iteration

Collect feedback from users and continuously improve the summarization algorithm based on their preferences and needs.

### Deployment

Deploy the audio summarizer as a standalone application or integrate it into existing podcast platforms.

### Monitoring and Maintenance

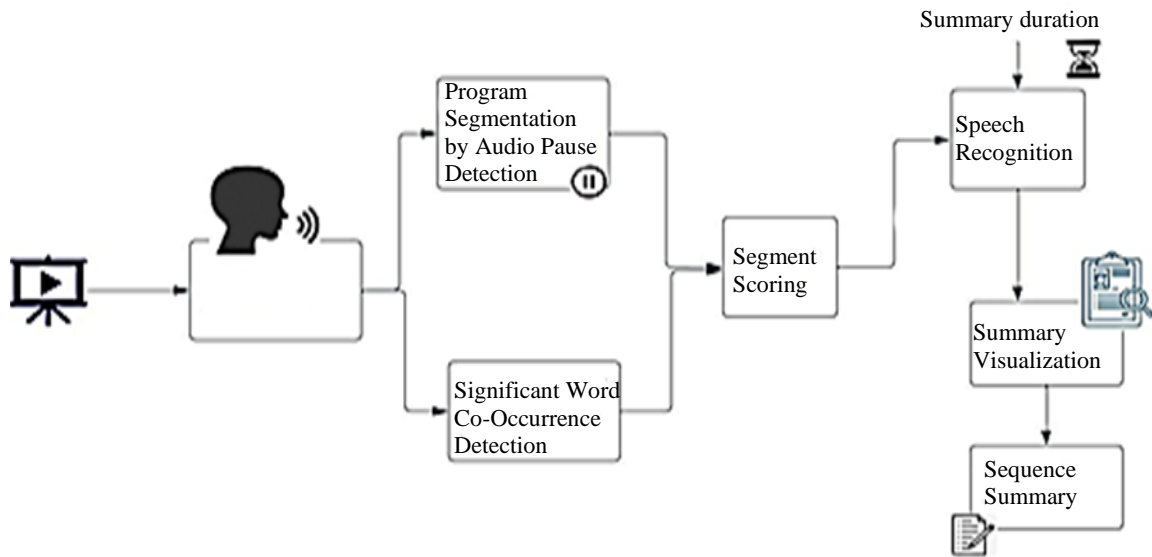
Regularly monitor the system's performance and make necessary updates to adapt to changes in podcast content and user requirements. The choice of specific NLP techniques, summarization algorithms, and TTS engines can vary depending on the project goals and resources. It is essential to fine-tune and optimize each step of the methodology to ensure that high-quality podcast summaries meet user expectations.

## FLOWCHART

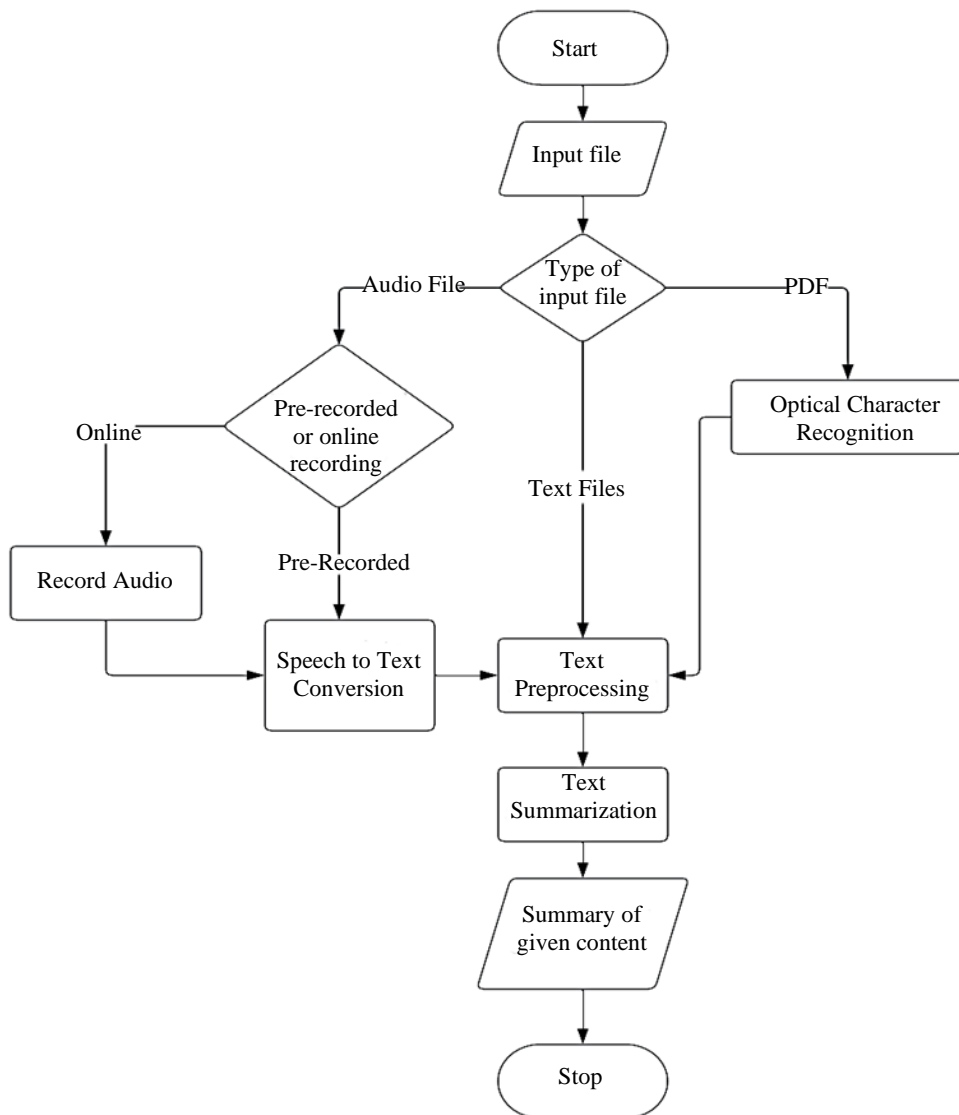
A flowchart for the audio summarization of podcasts visually outlines the process of extracting key points, creating concise summaries from podcast audio, and streamlining content for quick consumption [12]. It typically includes steps, such as audio input, transcription, key point extraction, and summary generation, as shown in Figure 2.

## RESULTS

The results of this research are presented in the Abstract. This includes a detailed description of the data collected and the results of the analysis. This section is organized systematically, starting with the most important findings [13]. Tables, charts, and graphs were used to clarify the results. Each figure and table are accompanied by a title explaining its content. The results are not disclosed, and their significance is discussed in the next section.



**Figure 1.** Proposed system of audio summarization of podcast.



**Figure 2.** Flow chart for audio summarization of podcasts.

### Creating the Main Screen

Figure 3 shows the main screen of the application. We created the first page, that is, the main page where the user can upload their request for summarizing the required podcast. By using Streamlit, which is an open-source framework, we created our application [14]. The first or main page requires podcast details, which are to be summarized. The user must upload a YouTube link to the podcast as well as which language the original podcast is generated.

### Working on the Application

After uploading the link of the podcast to be summarized and clicking on the transcribe button available at the bottom in Figure 4, the transcription process starts when the YouTube thumbnail of the podcast appears on the screen and simultaneously, the transcription process of the podcast occurs. For the conversion of speech-to-text (ASR), we used tools like GTTS (Google text-to-speech).

### Transcription and Generating Summary

Different processes, such as extractive and abstractive methods, have been used to generate a summary. Primarily, the text summary will be available on the screen, which can be read out by the user for better understanding [15]. If the podcast uploaded for summarization has been available in the Hindi language, then the text and audio summarization will both be in Hindi. Therefore, language selection is crucial for better understanding. As shown in Figure 5, the transcribed text is available.

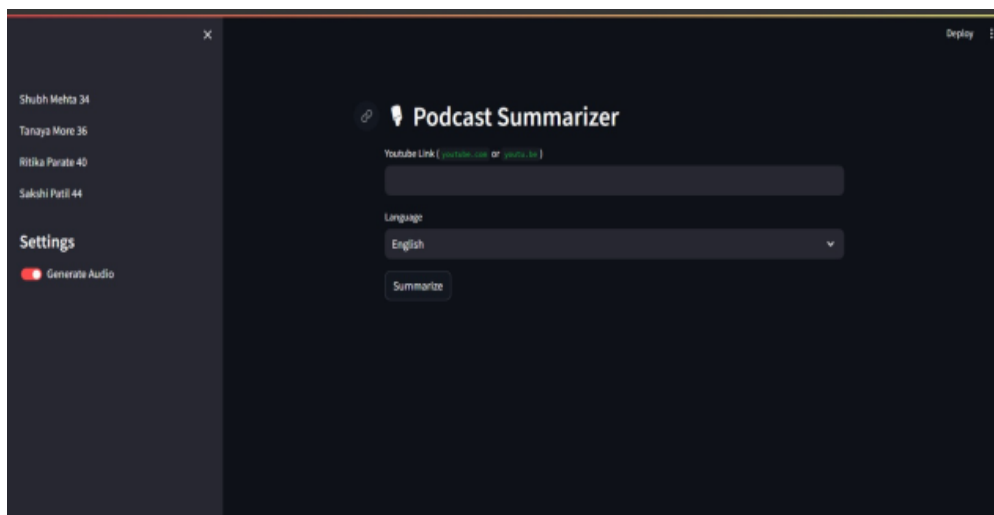


Figure 3. Main screen.

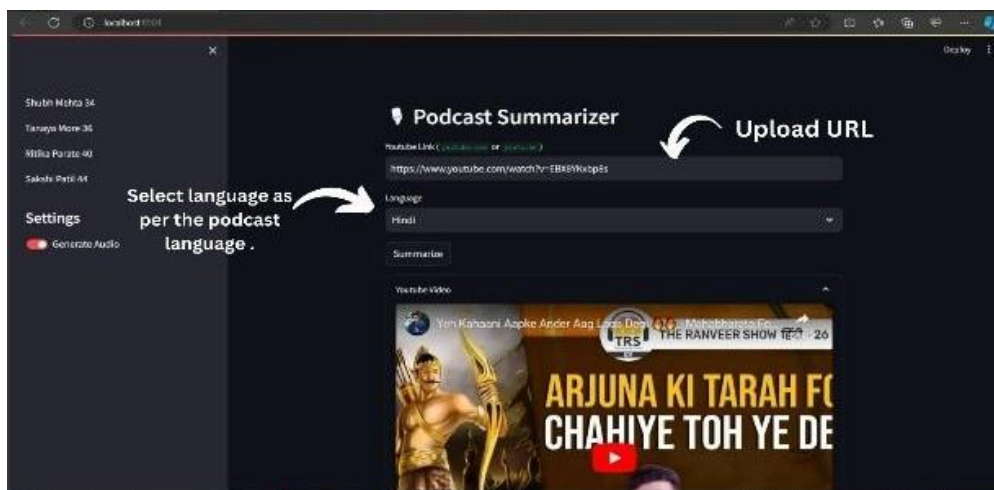


Figure 4. Uploaded podcast.

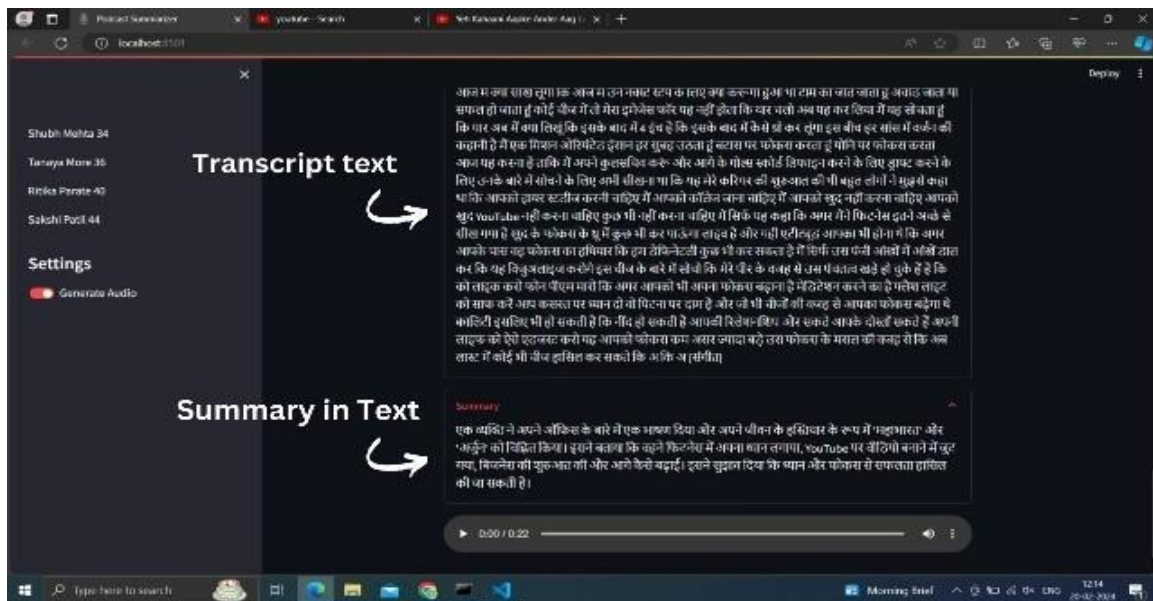


Figure 5. Transcript text and summary in text.

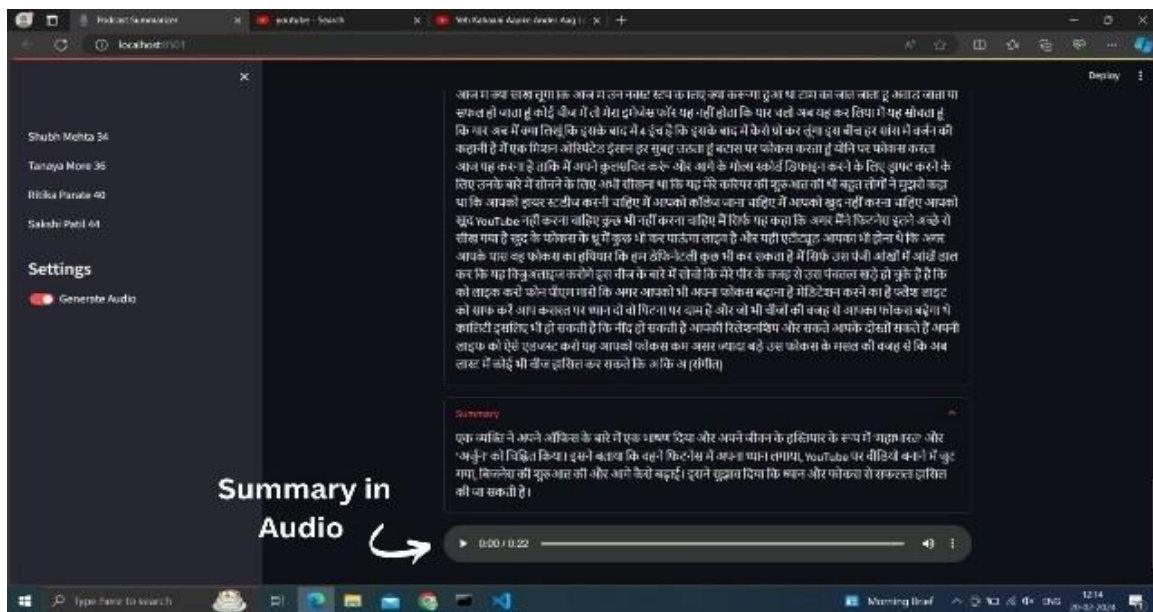


Figure 6. Summarized podcast of audio.

### Final Audio Summarization

The main aim of this application is to provide a summary of podcast episodes in audio. Here, the user will understand what the podcast consists of. For the conversation of speech-to-text, we used Eleven Labs, a software that specializes in developing natural-sounding speech synthesis using deep learning. The user will be able to listen to audio podcasts and obtain an idea of the podcast without wasting a lot of time, as shown in Figure 6.

### RESULT ANALYSIS

Table 2 examines the implications of the findings, their relevance to the research question, and their significance in the broader field. This chapter also discusses the unexpected results, possible limitations of the study, and suggestions for future research. The aim is to provide a better understanding of the results that contribute to the existing knowledge and to highlight the practical or theoretical significance of the findings.

**Table 2.** The test case of the proposed system.

ID	Test cases	Expected output	Actual output	Pass/fail
1	Upload a podcast audio file	Successful upload confirmation	The summarize button gets highlighted.	Pass
2	Initiate audio summarization	Summarized text or key points generated	summarized text or key points.	Pass
3	Test with different languages	Support summarization for podcasts in various languages	Summarized audio in selected language.	Pass
4	Verify handling of multiple speakers	Accurate summarization even with multiple speakers	Summarized text for multi-speaker podcast.	Pass
5	Test with podcasts of various durations	Accurate summarization for short and long podcasts	summarization for various durations.	Pass
6	Handle different audio formats	Summarization works for MP3, WAV, etc.	Only URL input works.	Fail
7	Check summarization length	Summary length within specified limits	The context length limit is 16385 tokens.	Fail

## CONCLUSION

In conclusion, audio summarization of podcasts is a valuable tool that enhances the podcast listening experience by providing concise and informative summaries of episodes. This technology can save time and help busy listeners discover content that aligns with their interests. As podcasting continues to grow in popularity, audio summarization algorithms are playing a crucial role in improving the accessibility and discoverability of podcast content. However, it is essential to ensure that the summarization process maintains the integrity and context of the original content, striking a balance between brevity and informativeness. The future of podcast audio summarization has great potential, and further research and development in this field can lead to even more sophisticated and user-friendly solutions.

## Future Scope

Audio summarization for podcasts is poised for the future of increased accuracy and efficiency with advancements in ASR and summarization. Imagine summaries that not only capture key points but also sentiment and speaker insights, allowing for interactive exploration and personalized recommendations. Real-time summaries and multilingual support could further enhance accessibility, whereas integration with podcast platforms would revolutionize content discovery.

## REFERENCES

- Rista A, Kadriu A. Automatic speech recognition: A comprehensive survey. *SEEU Rev.* 2020;15:86–112. DOI: 10.2478/seeur-2020-0019.
- Liya BS, Seenuvasan V, Sathya Prakash K, Siva B. Audio summarization in real time for podcast, speeches and audio books. *J Surv Fish Sci.* 2023;10:2618–25.
- Rachabathuni PK. A survey on abstractive summarization techniques. 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017. p. 762–5. DOI: 10.1109/ICICI.2017.8365239.
- Biswas PK, Iakubovich A. Extractive summarization of call transcripts. *IEEE Access.* 2022;10:119826–40. DOI: 10.1109/ACCESS.2022.3221404.
- Vanjari N, Pinjari S, Labde S, Patil P, Patil P, Sahitya A. Podcast summarization system. 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022. p. 379–82. DOI: 10.1109/ICAST55766.2022.10039668.
- Khanna D, Bhushan R, Goel K, Juneja S. SummarizeAI – Summarization of the Podcasts. Available at SSRN 4628657. *SSRN Electron J.* 2023. DOI: 10.2139/ssrn.4628657.
- Lee SW, Ralph TC, Jeong H. Fundamental building block for all-optical scalable quantum networks. *Phys Rev A.* 2019;100:052303. DOI: 10.1103/PhysRevA.100.052303.

8. Vartakavi A, PodSumm GA. Podcast Audio Summarization. arXiv Preprint ArXiv:2009.10315. 2020 Sep 22.
9. Zhang Y, Ma Q. Dual attention model for citation recommendation with analyses on explainability of attention mechanisms and qualitative experiments. *Comput Linguist.* 2022;48:403–70. DOI: 10.1162/coli\_a\_00438.
10. Mehta V, Deshpande T, Pandey R, Kandoi T, Nair S. Video transcript extraction and summarization using transfer learning. In: *International Conference on Information and Communication Technology for Competitive Strategies*. Springer Nature Singapore; 2022. p. 625–36.
11. Laban P, Ye E, Korlakunta S, Canny J, Hearst M. Newspod: Automatic and interactive news podcasts. In: *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 2022 Mar 22. p. 691–706. DOI: 10.1145/3490099.3511147.
12. Al-Fraihat D, Sharrab Y, Alzyoud F, Qahmash A, Tarawneh M, Maaita A. Speech recognition utilizing deep learning: A systematic review of the latest developments. *Hum-Centric Comput Inf Sci.* 2024;14.
13. Esmailzadeh S, Peh GX, Xu A. Neural Abstractive Text Summarization and Fake News Detection. arXiv Preprint ArXiv:1904.00788. 2019 Mar 24.
14. Nallapati R, Zhou B, Gulcehre C, Xiang B. Abstractive Text Summarization Using Sequence-to-Sequence Rnns and Beyond. arXiv Preprint ArXiv:1602.06023. 2016 Feb 19.
15. El-Kassas WS, Salama CR, Rafea AA, Mohamed HK. Automatic text summarization: A comprehensive survey. *Expert Syst Appl.* 2021;165:113679. DOI: 10.1016/j.eswa.2020.113679.