

# An Intelligent System for Automated Meeting Notes Generation Using Recognition and NLP.

Angel Sandeep Mahadik<sup>1</sup> and Prof. Mrs. Vidya T. Thakare<sup>2</sup>.

Student, Department of Computer Engineering<sup>1</sup>

Co-ordinator, Department of Computer Engineering<sup>2</sup>

JSPM's Rajarshi Shahu College of Engineering, Polytechnic, Tathawade, Pune, Maharashtra, India

**Abstract:** Meetings are an integral part of organizational or institutional-based decision making, project implementation, or dissemination of know-how. Often, manual note writing at such meetings tends to be not so effective and may lead to inaccurate note writing. Therefore, to overcome such drawbacks, this paper proposes a Smart Meeting Notes Generator using speech-to-text processing, Natural Language Processing, and AI-based Summary Generation. The above notwithstanding, other benefits of adopting such a system are discussed under the next heading.

The system analyses the recorded audio of the meeting in order to produce an accurate text version of the proceedings, from which the brief summaries and key points of the discussions can be extracted. Furthermore, the system extracts significant decisions that emerged from the discussions in addition to the action items that need follow-up attention for proper documentation. The system is user-friendly and supports multi-device functionality so that the user can easily access the minutes of the meetings from the desktop or mobile versions of the software. The Smart Meeting Notes Generator system boosts productivity while enhancing documentation efficiency in meetings.

**Keywords:** Smart Meeting Notes Generator, Speech-to-Text, Natural Language Processing, Automatic Text Summarization, Meeting Documentation, Audio Processing, Action Item Extraction, AI-Based Systems

## I. INTRODUCTION

Meetings are a fundamental part of the organizational environment, being a major platform for carrying out decisions, organizing projects, strategic planning, as well as the dissemination of knowledge. In a business setting or learning centers, meetings are organized on a regular basis in order to discuss the ongoing developments, delegate tasks, as well as capture significant decisions. Meeting success is mostly dependent on the level of accuracy in recording the proceedings.

Although the role of documenting the minutes of the meeting is important, the process through manual note-taking is quite difficult. The fact is that manual recording is very much subject to the level of the person who is recording the minutes. In fact, factors such as the pace of the proceedings, the number of persons speaking simultaneously, technical vocabulary, and the length of the proceedings may result in the minutes being inaccurate or incomplete. Many times, important decisions or actions may go unnoticed.

Recent progress in AI, speech-to-text processing, and NLP has facilitated the development of intelligent systems that can process audio data and produce meaningful textual outputs. Speech recognition technologies convert spoken languages into written text; natural language processing techniques further process the written text to enable summarization, keyword extraction, and the identification of key statements. This lays a platform for automating the structuring of meeting notes and reducing the manual effort required.

The advantages of automated meeting note generation systems are that they enhance the accuracy of documentation, increasing accessibility and efficiency. By distilling the key discussion points, decisions, and action items, this system enables the participants to review meeting outcomes quickly without having to listen to long recordings. Digital



systems also allow easy storage, search, and sharing of meeting notes across platforms, hence supporting effective collaboration in professional and academic settings.

This paper offers a solution for the Smart Meeting Notes Generator that can combine speech-to-text processing, NLP, and AI methods for summarization in order to generate meeting minutes from audio recordings automatically. The proposed system can take audio from meetings and transcribe it into text, summarize meetings conducted, and identify important decisions and actions that need to be taken. The proposed system is very user-friendly and can work on multiple devices in order to provide access to meeting notes through desktop and mobile support. Therefore, the proposed system can increase productivity and accuracy in managing meetings. The remainder of this paper is divided into the following sections: Section II will discuss related work in the field of automated meeting documentation and speech processing. Section III will detail our proposed method and system architecture. Section IV will present results and discussion, followed by conclusions and future scope in Section V.

## **II. PROBLEM STATEMENT**

They are commonly employed in organizational communication, decision-making, and coordination. Nonetheless, the process of accurately recording the discussions that take place in the meeting is a challenge. The traditional method of note-taking in the meeting is inefficient and prone to human error, especially in long discussions or in meetings that have many participants. This could result in the recording of important details such as the main points, decisions, and action items.

While audio recordings can be relied upon for reference, it is time-consuming to refer to the entire audio for making decisions or for further action. In particular, transcribing these recordings into text has been proven to be a time-consuming task that also requires expertise. While tools for transcribing may be available, tools for summarization or structured information extraction related to making decisions or action items may not be so effective. Thus, there is an obvious requirement for an accurate system that can automatically transcribe important meeting audio into structured meeting notes.

## **III. OBJECTIVES**

The prime task of this project is to design and develop a Smart Meeting Notes Generator system automatically documenting meetings through speech-to-text processing, natural language processing, and AI techniques. The specific objectives of this system include the following:

- To translate the live or pre-recorded audio of a meeting into an accurate and readable form of text by the use of the "speech to text"
- To enable the use of sophisticated natural language processing methods for writing concise summaries of the discussions at lengthy meetings.
- To automatically identify and tag key action items, decisions, and deadlines from the discussions in meetings.
- For providing the facility to export in multiple formats, such as PDF, Word documents, and Emails, in order to share.
- For secure and organized storages of the generated meeting notes in a data base with efficient search capabilities to easily look up past discussions.
- For real-time access and collaboration, to allow several users to view and edit meeting notes simultaneously.
- To enable transcription and summarization functionalities for varied languages, catering to varied teams worldwide.
- To ensure the confidentiality of meeting recordings, transcriptions, or summaries by encrypting them.

## **BRIEF EXPLANATION**

The purpose of the requirement to translate the audio recording of the meetings to written form is to ensure accurate recording of the discussions. By implementing the speech recognition technique through the system, the verbal interaction that takes place in the process of the meeting is put down in written form.



- The goal of creating concise summaries is to offer the user an understandable and rapid insight into the conversation that takes place in the meeting. This means that instead of having to evaluate the lengthy transcript or recording, they will be able to understand the topic in a short span of time.
- It becomes vitally important to identify and extract key decisions and action items from such meetings to effectively follow up on them. This goal helps to emphasize key results of meetings so that members are able to monitor them without any confusion.
- The reliance on manual note-taking is relevant to the limitation that comes with human error, fatigue, and divided attention in a meeting.
- Finally, having an easily usable and multi-device compatible platform enables the consideration that meeting minutes can be accessed and shared efficiently across different devices, including desktop and mobile devices.

#### **IV. LITURATURE REVIEW**

There has been significant research and implementation of speech recognition and natural language processing technologies in automated documentation systems. Microsoft Azure Speech Services (2024) is one such cloud speech-to-text platform that has the capability to transcribe audio, including live and pre-recorded audio, into text accurately. The platform has support for various languages and speaker identification. There could be potential delays due to dependence on the cloud and the internet.

There has been research centered on utilizing natural language processing-based summarization techniques for extracting important information from large bodies of text. Numerous studies have been done on extractive and abstractive techniques for enhancing summaries by making them more succinct and clearer. Even though these techniques work well for condensing text and pointing out important information, it still has challenges regarding action items in conversations.

Another common speech recognition system that has extensively been used is the IBM Watson Speech to Text. This is offered by IBM (2023) as a business-oriented speech recognition tool with support for customizable language models. Although it has flexibility in its usage, its performance may depend on audio file quality and terminology used in the applications.

Google Speech-to-Text uses deep learning-based models for accurate and scalable speech recognition in a host of applications. It supports a variety of languages and audio formats, which is quite useful in large-scale deployments. However, similar to other cloud solutions, the complexity of integration and cost considerations may impose limitations on its applicability in some scenarios.

The review clearly indicates that the available solutions using speech recognition and NLP-based summarization form firm bases for the creation of automated meeting documentation. Most of these approaches propose single components, either transcription or summarization, without providing a complete integrated system for automatic generation of meeting notes. In this regard, there is a need for an intelligent end-to-end integrated system that uses speech and NLP techniques in generating structured and actionable meeting notes.

#### **V. SYSTEM ARCHITECTURE**

The architecture of the proposed system for automatically generating meeting notes has a modular client-server architecture along with artificial intelligence. As can be seen in Figure 1, the system has been designed for processing audio recordings of meetings and automatically generating structured meeting notes through natural language processing.

It begins with the user end, where audio files of meetings are uploaded using a web-based interface. The uploaded files are then handled by Multi middleware to perform secure uploads of files and transfer audio to the file storage module. The audio file stored in this module is then further processed using FFmpeg to convert it to a Standard MP3 format. This ensures compatibility with services of speech recognition.



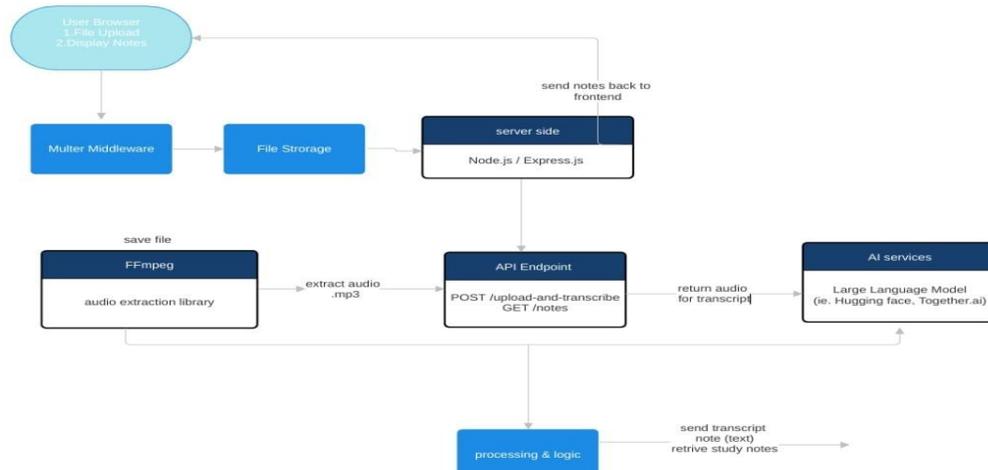


Figure I – System Architecture

The backend server design uses Node.js and the Express framework. This server has API routes that enable the management of the audio files being uploaded as well as the retrieval of the generated notes. These processes involve the transfer of the audio files to the artificial intelligence language services. These services enable the transcription of the audio files to written text. This written text is then analysed by the backend server to generate the final notes. The generated notes are then sent back to the frontend and displayed to the user via the browser interface. Such architecture ensures proper interaction among system components, scalability, and trustworthy automation of meeting documentation.

## VI. SOFTWARE & HARDWARE REQUIREMENT

### 1. Software Requirements

The proposed system is developed using modern web and artificial intelligence technologies. The frontend interface uses HTML, JavaScript, and CSS to create a responsive and user-friendly environment for meeting audio uploads and generated notes. Node.js and Express.js are employed in the backend to handle server-side logic and API communication. Multer middleware is used to handle file uploading, while FFmpeg is used for extracting and converting audio data. In addition, this system integrates the use of external APIs for speech transcription and meeting notes generation. It also uses MongoDB for structured data storage, where, depending on the nature of deployment, data might be stored using local or cloud-based storage solutions. Finally, its AI components consist of speech-to-text and text summarization models like those provided by various GPT-based services or platforms such as Hugging Face.

### 2. Hardware Requirements

At the client side, any modern laptop or smartphone with a microphone and recent web browsers like Google Chrome, Microsoft Edge, or Mozilla Firefox will work. On the server side, this would require at least 8 GB RAM, 50–100 GB SSD storage to serve backend services and store files, with API execution. Ubuntu or Windows could be an operating system for this server. The stable internet connection requires at least 100 Mbps bandwidth. However, in those deployments where the local execution of AI models is involved, higher computational resources such as 16–32 GB RAM and approximately 500 GB SSD storage will be required for efficient model inference and data processing.

## VII. RESULT AND DISCUSSIONS

The system was implemented and tested, and the results are indicated and well analyzed that it can showcase the generated notes with giving accurate and well analyzed with structured and definition. It displays the generated notes and all the features.



TABLE I – PERFORMANCE CHART

Sr.No	Performance	Values	Units
01.	Login Functionality Accuracy	99.7	Percentage (%)
02.	Sign-In Page Success Rate	98.3	Percentage (%)
03.	Audio Upload Processing Time	2.8	Seconds (s)
04.	Speech-to-Text Transcription Time	4.1	Seconds (s)
05.	Summary Generation Time	3.6	Seconds (s)
06.	Meeting Notes Generation Time	7.5	Seconds (s)
07.	System Response Time	1.9	Seconds (s)
08.	Overall System Accuracy	97.8	Percentage (%)

This Table 1, represents the performance evaluation of the proposed system. The results indicate high accuracy for core functionalities such as login and sign-in operations, along with acceptable processing times for transcription and summarization. These results demonstrate the efficiency and reliability of the system for automated meeting notes generation.

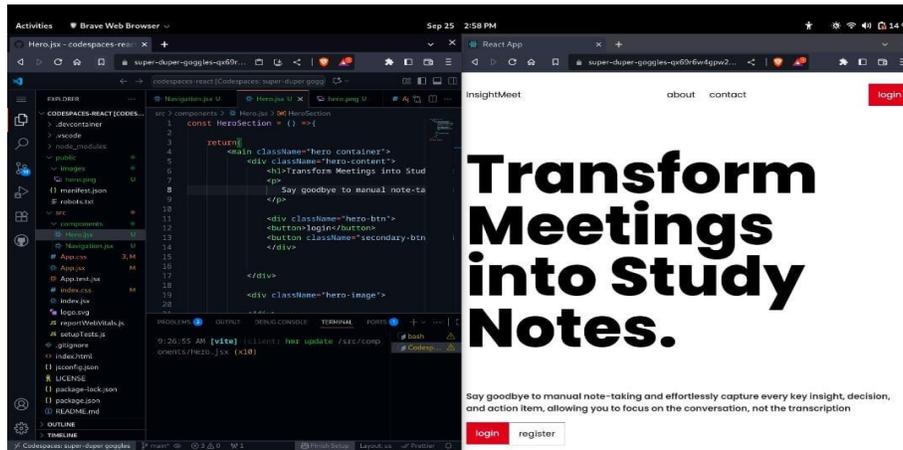


FIGURE II – WELCOME PAGE OF WEBSITE



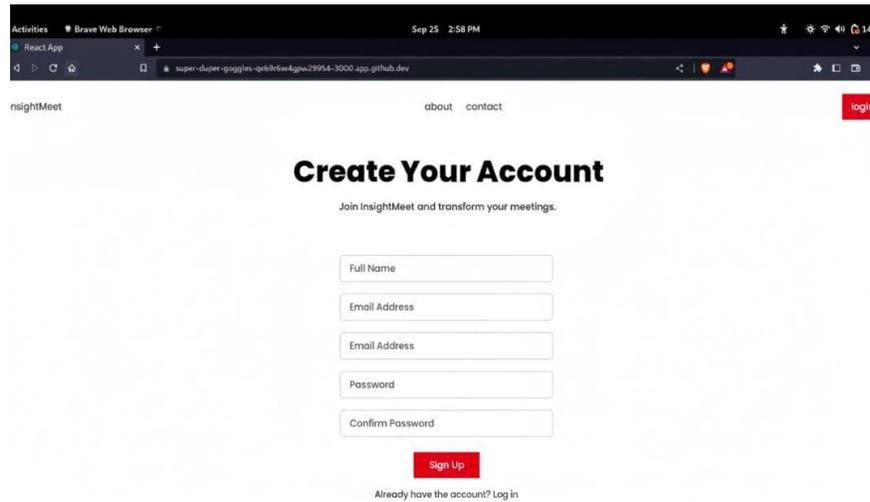
## Sign in to your Account



Don't have an account? [Register](#)

FIGURE III – SIGN IN \ LOG IN PAGE UI





**FIGURE IV – REGISTRATION PAGE UI**

### VIII. CONCLUSION

In this paper, an intelligent system for automatically generating meeting notes from speech recognition and natural language processing was discussed. The proposed solution has been put forward as it tackles the shortcomings of having human-captured meeting notes by offering an automation process for transcribing, summarizing, and arranging meeting discussions. With web interfaces, processing services, and language models leveraged via AI, this solution enables an efficient process for structured meeting notes from meeting audio recordings.

Results from the performance evaluation indicate that the developed system is capable of delivering high accuracy in the key functionalities related to authenticating the users and the transcription processes. Additionally, the developed system is capable of processing audio and performing the summary functions within acceptable amounts of time. The designed modular architecture is capable of ensuring ease-of-use and functionality. In conclusion, the developed solution is capable of boosting productivity and accessibility to the information in the meeting.

### ACKNOWLEDGMENT

Author Angel Sandeep Mahadik would like to extend their sincere acknowledgments to Prof. Mrs. Vidya T. Thakare, Coordinator, Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Polytechnic, Tathawade, Pune, Maharashtra, India, for her constant support, suggestions, and encouragement in the preparation of this project.

The author is also grateful to the faculty and staff of the Department of Computer Engineering of JSPM's Rajarshi Shahu College of Engineering, Polytechnic, for the requisite infrastructure and support that made the successful completion of this task possible.

Secondly, the appreciation is extended to peers, colleagues, and others for their valuable suggestions and help throughout the project stages.

### REFERENCES

- [1]. Google Cloud, "Google Speech-to-Text API Documentation," 2024. [Online]. Available: <https://cloud.google.com/speech-to-text>. [Accessed: Jan. 10, 2026].
- [2]. IBM, "IBM Watson Speech to Text," 2023. [Online]. Available: <https://www.ibm.com/cloud/watson-speech-to-text>. [Accessed: Jan. 10, 2026].
- [3]. Microsoft Azure, "Azure Speech Services Documentation," 2024. [Online]. Available: <https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/>. [Accessed: Jan. 10, 2026].



- [4]. D. Jurafsky and J. H. Martin, \*Speech and Language Processing\*, 3rd ed., Pearson, 2023.
- [5]. S. Young et al., "The HTK Book," Cambridge University Engineering Department, 2023. [Online]. Available: <http://htk.eng.cam.ac.uk/>.
- [6]. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
- [7]. A. Vaswani et al., "Attention Is All You Need," \*Advances in Neural Information Processing Systems\*, vol. 30, pp. 5998–6008, 2017.
- [8]. C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," \*Journal of Machine Learning Research\*, vol. 21, no. 140, pp. 1–67, 2020

