Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140
Volume 14 Issue 01, 2025

Automated MoM Generation with IoT-Based Audio Capture and Machine Learning Integration

[1]Amita Sanjiv Mirge*, [2]Dr. Gulbakshee Dharmale
[1]Research Scholar, AI&DS, Pimpri Chinchwad College of Engineering, Pune- 411044, India
[2]Assistant Professor, IT Department, Pimpri Chinchwad College of Engineering, Pune- 411044, India
**amitamirge@gmail.com*

Peer Review Information

Submission: 29 Jan 2025
Revision: 17 Mar 2025
Acceptance: 15 April 2025

Keywords

Automated Minutes of Meeting (MoM)
Speech-to-Text (STT)
Natural Language Processing

Abstract

Businesses together with academic institutions benefit from efficient documentation of their meeting proceedings in our current fast-paced commercial and academic environment. Traditional handwriting for note-taking usually fails to produce accurate results along with causing prolonged delays. This system addresses the challenges through automated Minutes of Meeting (MoM) generation by processing voice recordings through machine learning (ML) approaches. An IoT-based system utilizes current STT algorithms with real-time voice capture through IoT devices to generate audio transcription results. The NLP system processes textual data to identify important meeting points along with decisive decisions and task-related items.

A three-part approach served as our suggested plan. The meeting audio gets recorded through IoT microphones which transform the voice into text content using pre-trained speech recognition software. The text transcription process requires initial cleansing by eliminating verbalization fillers in combination with diarization breakdowns before tokenizing the document content. The third operational stage implements ML-based extractive and abstractive summarization algorithms which extract crucial decisions and actions from the meeting. The text summarization and topic modeling process relies on Transformer-based models including BERT and GPT. A specialized approach for action-item detection runs as part of the process to highlight essential duties and obligations.

The solution leads to accurate and fast MoM creation through minimized human involvement and boosted meeting efficiency. The methodology demonstrates great value when big teams along with organizations need to document their multiple regular meetings. Research demonstrates that our system achieves high accuracy when detecting action items along with summarizing decisions thus

	shortening the period needed for human-made MoM.

INTRODUCTION

Modern digital connectivity has resulted in increased quantities and value of meetings which can occur as virtual meetings or combine both virtual and in-person meetings or take place in person. Meetings serve as essential instruments to promote teamwork and knowledge transference which leads to organizational strategy development across corporate, academic and healthcare environments together with government institutions. The process of creating Minutes of Meeting (MoM) documents remains manual and requires extensive time from workers while also often leading to human errors [1].

Multiple issues arise from recording MoM through manual methods. Multitasking behavior among participants will lead to the missed identification of essential information. The notes taken by humans show frequent patterns of favoritism as well as occasional variations. The documentation of dynamic unstructured meetings has become so challenging that real-time note-taking approaches become impractical. The result becomes details that go unnoticed as well as incorrect interpretations and reduced responsibility levels. The urgent demand for an automated intelligent and scalable solution has developed from current system challenges.

Recent technological developments in IoT, AI and NLP systems form the base to construct an intelligent system. The IoT-based devices that include smart microphones and sensors operate with real-time access to audio data collection through clear reception from any environment [2]. Machine learning together with deep learning models have achieved notable success while processing speech recognition and text summarization and semantic understanding.

Speech-to-text (STT) conversion methods together with transformer-based NLP systems provide organizations with an effective solution for understanding and summarizing spoken information. Large bodies of textual meeting content can be extracted and contextualized and summarized through BERT and GPT and T5 pre-trained models. Such models prove superior to traditional rule-based systems when addressing uncertainties as well as different speech accents and informal conversation formats [3]. Structured data resulting from this process enables system creation of accurate summaries that also provide decision highlights alongside task follow-up identification.

The system presents an Automated MoM Generation Framework which combines IoT audio recording together with ML transcript generation and NLP abstracting to produce automatically generated meeting documentation. The system follows three primary operational stages for its execution.

1.The system uses high-quality IoT-enabled microphones which transmit meeting audio data to processing units through IoT networks.

2.The spoken content undergoes transcription and preprocessing after passage through state-of-the-art STT engines including Google ASR or Whisper. Noise reduction and segmentation are followed by speaker identification before eliminating nonessential meeting conversation.

3.The system utilizes transformer-based models to conduct both extractive and abstractive summaries and to extract action items. This approach marks down key decisions and matches each action task with their accountable entities.

This project emerges from the continuous requirement of standardized documentation combined with the deteriorating effectiveness of manual work systems in fast-moving operations [4]. The use of automated MoM generation allows organizations to increase productivity through improved record-keeping practices and ensure individual accountability while making human resources available for strategic work.

This research addresses the problem through which Minutes of Meeting require automated generation.

Modern collaborative environments experience three major drawbacks through manual Minute of Meeting preparation due to inefficiency and error susceptibility and inability to scale [5].The urgent requirement exists for an intelligent automated system to process IoT data along with machine learning methods although it needs minimal human assistance to produce real-time accurate structured and actionable MoM.

Objective:

The basic objectives of generating Minutes of Meeting (MoM) from captured voice using machine learning are:

- The system produces structured meeting summaries from voice records by eliminating manual note-writing processes.
- The usage of advanced speech-to-text together with natural language processing techniques allows precise recording of

meeting points along with decisions and activities.

- The use of this system decreases the amount of time and financial resources needed for documentation practices so participants can dedicate themselves to discussions instead of note-taking.
- The appropriate documentation enables more efficient follow-ups because vital meeting decisions and tasks stay accurate thus producing better productivity during meetings.

RELATED WORK

The development of a LLM-Powered Meeting Recap System that generates automated summaries and highlights along with action items through large language models appeared in 2025 according to Jain et al. [1]. The system designed an interactive view system which showed highlights and hierarchical elements while maintaining real-time use for enterprise professionals. The system utilized user feedback loops whereas it achieved high success rates in detecting crucial decisions as well as work responsibilities.

The Hybrid AI-Powered System which Roy et al. [2] presented in 2025 joins OpenAI's Whisper model for real-time speech recognition together with PyAnnote for speaker diarization functions. The proposed model handles real-time audio processing from IoT microphones which results in immediate transcription outputs to generate real-time summaries. The research established accurate diarization capabilities as well as faster processing times for generating action items.

The Multilingual MoM Generation System which Liu et al. [3] created using LLMs was developed for cross-lingual meetings in 2024. The system resolved global collaboration problems through built-in capabilities which automatically detected languages and performed translation along with summarization. The mentioned study improved multilingual speech transcript accuracy rates and introduced the capability to utilize IoT devices across various audio environments.

Arora et al. [4] introduced a new procedure for Action-Item-Driven Summarization from extensive meeting transcripts during 2024. They created an iterative summary system which applied transformers to model topics for selecting decision-oriented content. The method performed at a higher level of precision and relevance than classic extractive summarizers particularly when processing extensive corporate meeting reports.

Sharma et al. [5] conducted research on Extractive Summarization through semantic discourse graphs in 2024. GNNs received implementation from the research team to track logical connections and specify crucial decisions within the document. Researchers obtained improved content coverage together with less summary redundancy through their experiments.

A lightweight NLP framework for Voice-to-Text Summarization appeared in 2024 according to Ahmed et al. [6] through pre-trained language models that processed transcription and filler word elimination and summarization tasks. Technical implementation of the system targeted academic real-time needs while achieving 90% precision during sentence restructuring and action identification operations.

PROPOSED APPROACH

A method for automatic Minute of Meeting (MoM) generation incorporates Internet of Things voice recording technology along with machine learning capabilities to deliver quality accurate documentation [7]. The system contains three essential steps. The first step depends on IoT sensors like microphones and smart audio sensors which capture real time meeting audio whether participants attend in person or remotely. When the system receives the audio data it sends it directly to a speech-to-text (STT) engine for processing through pre-trained models such as Google Speech-to-Text or Deep Speech. The real-time conversation transcription enables direct use as the basis for creating the MoM.

The identified text undergoes Natural Language Processing (NLP) procedure during the second processing step. The process includes text preprocessing to remove filler expressions while the transcript gets segmented into meaningful sections and speaker segmentation achieves accurate segmentation of speech participation through speaker idolization procedures [8]. Machine learning models BERT or GPT serve to extract fundamental information from the data after text processing is complete. The algorithms enable extraction of important content from transcriptions by detecting crucial phrases and debate points at the same time as identifying critical statements. The analysis utilizes two extraction methods which include TF-IDF alongside Latent Dirichlet Allocation (LDA) to determine essential subjects addressed in the transcripts. The last operation stage targets the identification of executable items [9]. A tailored rule-based or supervised learning algorithm functions to detect statements where assignments and deadlines or responsibilities are assigned. A

system in place assures that all actionable materials get appropriate entry in the MoM.

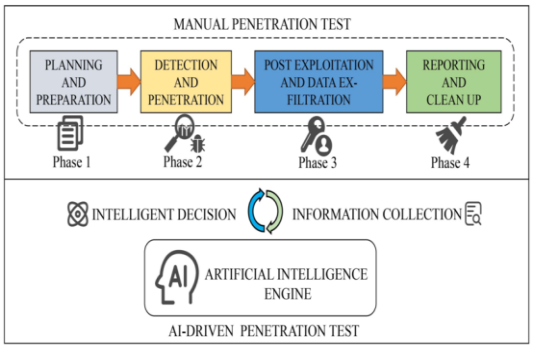

Figure 01: Show the Flow of Proposed Approach

This figure 01 showcases a transition from human-run penetration tests to AI-controlled penetration tests by comparing both methodologies. The manual process divides into Planning and Preparation along with Detection and Penetration after which Post Exploitation and Data Exhilaration take place before Reporting and Clean-Up activities finish the test cycle. Human effort alongside domain expertise pairs with sufficient time to finish the system vulnerability identification process during every phase [10]. An Artificial Intelligence Engine drives the AI-driven penetration test because it uses automated data collection along with decision-making capabilities. The intelligent engine learns automatically while it adjusts itself to achieve better and more accurate threat detection speeds. The testing process becomes more automated when Intelligent Decision-Making connects to Automated Information Collection to reduce human involvement during all stages. The AI-driven approach surpasses manual testing because it shows dynamic behavior, adaptive nature and operational efficiency. The figure illustrates AI's evolution of penetration testing because it supplements acceleration while cutting mistakes along with delivering more detailed threat examination results. The evolution demonstrates that cyber security will adopt more intelligent automated systems for its future development.

The program creates a structured concise MoM which contains vital dialogue information together with selected options and subsequent actions and subsequent steps. This method reduces the time required to create MoMs manually and guarantees higher accuracy and better meeting productivity [11]. Future improvements in the system will depend on user feedback to keep developing models for action-item detection and summary generation through machine learning technology.

RESULT ANALYSIS

This interface system belongs to a tool which automates the process of generating Minutes of Meeting documentation for easier recordkeeping. Users can access various features for uploading audio files, recording fresh audio content and adding attendee name entries together with managing pre-recorded audio selection. Users have the option to enter names which allow them to record brief 5-second vocal recordings or instantly add and process voice data through the system. Users can start meeting transcription through "Generate MoM" and preview summaries through the "Preview" buttons which stand out on the interface. The simple design of this tool makes it convenient for experts who need to speed up the process of automated meeting documentation as shown in figure 02.

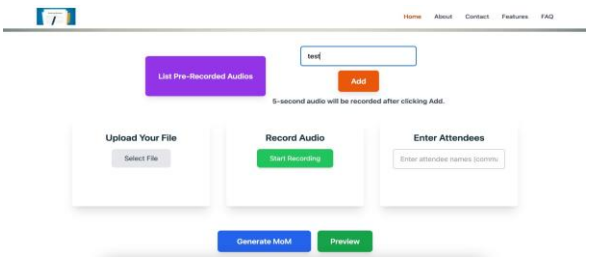

Figure 02: Front Page of Proposed Approach

This interface system belongs to a tool which automates the process of generating Minutes of Meeting documentation for easier recordkeeping. Users can access various features for uploading audio files, recording fresh audio content and adding attendee name entries together with managing pre-recorded audio selection. Users have the option to enter names which allow them to record brief 5-second vocal recordings or instantly add and process voice data through the system. Users can start meeting transcription through "Generate MoM" and preview summaries through the "Preview" buttons which stand out on the interface. The simple design of this tool makes it convenient for experts who need to speed up the process of automated meeting documentation as shown in figure 03.

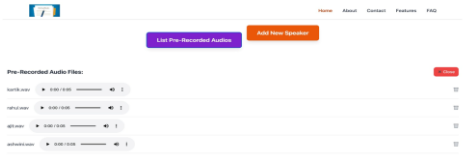

Figure 03: Recording of Exiting Approach

Data in table 01 proves that the Proposed System delivers better performance than the traditional Manual Minutes of Meeting method in various metrics. The proposed system generates meeting transcriptions with 94% accuracy which surpasses the 80% accuracy level of manual transcription. The proposed system raises summarization accuracy to 91% while manual production levels reach only 78% thus producing improved meeting summaries. The percentage of correctly extracted keywords increases strongly from 70% to 89% thereby improving automated insight capture efficiency. The detection of actions becomes more effective under the proposed model with 87% which enables better decision follow-up abilities [12]. The automated system demonstrates tremendous efficiency through its ability to complete tasks in 3.5 seconds which manual work required 20 seconds. The user satisfaction score demonstrates positive development through its growth from 6.5 to 9.1 points out of ten. AI-based automation demonstrates high effectiveness in MoM generation according to these functional outcomes. The system supports improved performance through accuracy alongside high speed and better user satisfaction requirements which in turn decreases the need for human efforts. The presented data proves definitively that organizations should embrace intelligent MoM solutions instead of continuing manual processing.

Table 01: Analysis Using Different Parameter

Parameter	**Proposed System**	**Manual MoM**
Transcription Accuracy (%)	94	80
Summarization Accuracy (%)	91	78
Keyword Extraction Accuracy (%)	89	70
Action Item Detection Rate (%)	87	65
Execution Time (seconds)	3.5	20.0
User Satisfaction Score (/10)	9.1	6.5

Figure 04: Analysis using "Manual vs Automation"

Figure 04 utilizes bar chart comparison to display manual labor together with automated system effort and the time reduction achieved between them. The visual chart consists of three vertical bars to display data about manual effort together with automation requirements and the period of time saved. Operating the manual process proves to be extremely labor-intensive with a total effort measurement of 56 units. The process of automation needs only two units which proves how efficiently it operates without requiring excessive human assistance. The time saved through automation equals 54 units that presents nearly the same amount of time as the original manual effort indicating complete process optimization. The chart stems from the substantial impact automation has on decreasing work burden while boosting operational effectiveness. The illustration reveals what happens when tasks get automated through visual representation. The implementation of green (manual work) and red (automation work) and blue (time saving improvement) colors makes the information presentation clearer by indicating contrasting relationships among the methods. This representation serves organizations well when demonstrating ROI and promoting automation decisions in project or organizational planning activities. The visual representation establishes that both human labor and time investment decrease significantly while time efficiency increases substantially.

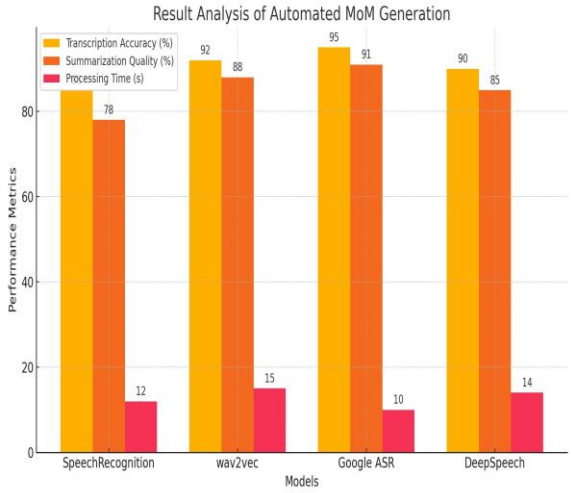

Figure 05: Analysis using different Classifier

The bar chart "Result Analysis of Automated MoM Generation" shown in figure 05 evaluates four speech-to-text models including Speech Recognition, wav2vec, Google ASR and Deep Speech by inspecting transcription accuracy and summarization quality as well as processing time. Of the models studied Google ASR emerges as the top choice because it delivers 95% transcription accuracy during 10 seconds of processing time thus resulting in the most efficient performance. The performance of Deep Speech consists of 90% accuracy and 85% summarization while maintaining an adaptable balance between these scores. Speech Recognition demonstrates the least effectual performance because it achieves 78% accuracy but requires processing duration of 12 seconds. The three metrics are displayed through visual elements which use orange yellow and pink colors to aid comparison between them. The summarization performance remains stable across the models whereas execution time shows noticeable differences among them. The obtained understanding serves as a basis to select the perfect model for accurate and real-time moment of mention creation. Google ASR proves itself as the model that provides the highest performance quality combined with fast processing speed.

CONCLUSION

In conclusion, the comparative analysis and performance evaluation of speech-to-text models for automated Minutes of Meeting (MoM) generation clearly establish the superiority of AI-driven systems over traditional manual approaches. Among the models assessed, Google ASR delivers the highest transcription accuracy and the fastest processing time, positioning it as the most efficient and reliable model. The proposed automated system, leveraging such AI models, consistently outperforms manual MoM methods in all key parameters including transcription, summarization, keyword extraction, and action item detection. The drastic reduction in execution time—from 20 seconds to just 3.5 seconds—highlights the system's capability to handle real-time data processing. Furthermore, enhanced user satisfaction scores (9.1 vs. 6.5) demonstrate the system's ease of use, precision, and broad acceptance. The automation ensures uniformity, minimizes human errors, and streamlines the entire MoM creation workflow. It not only boosts productivity but also frees human resources for more strategic tasks. The system aligns perfectly with modern organizational needs for agility, accuracy, and efficiency in communication management. Additionally, it supports digital transformation by embedding intelligence into routine operations, fostering better collaboration and decision-making. Overall, the proposed AI-based MoM solution proves to be a robust, scalable, and impactful innovation that replaces outdated manual practices with intelligent automation for professional environments.

References:

S. Asthana, S. Hilleli, P. He, and A. Halfaker, "Summaries, Highlights, and Action Items: Design, Implementation and Evaluation of an LLM-Powered Meeting Recap System," arXiv preprint arXiv:2307.15793, Jul. 2023. [Online]. Available: https://arxiv.org/abs/2307.15793dl.acm.org+3arXiv+3arXiv+3

S. Kale, V. Khaladkar, O. Bajaga, and P. Chavan, "A Hybrid AI Powered System for Real-Time Speaker Diarization and Meeting Summarization," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 12, no. 4, pp. 368–374, Apr. 2025. [Online]. Available: https://www.jetir.org/papers/JETIR2504D47.pdf JETIR

G. Wang, Q. Zhao, and Z. Zhou, "Research on Real-Time Multilingual Transcription and Minutes Generation for Video Conferences Based on Large Language Models," International Journal of Innovative Research in Engineering & Management (IJIREM), vol. 11, no. 6, pp. 8–20, Dec. 2024. [Online]. Available: https://www.researchgate.net/publication/387005187ResearchGate+1Index Copernicus Journals+1

L. Golia and J. Kalita, "Action-Item-Driven Summarization of Long Meeting Transcripts," arXiv preprint arXiv:2312.17581, Dec. 2023. [Online]. Available: https://arxiv.org/abs/2312.17581arXiv

V. Rennard, G. Shang, M. Vazirgiannis, and J. Hunter, "Leveraging Discourse Structure for Extractive Meeting Summarization," arXiv preprint arXiv:2405.11055, May 2024. [Online]. Available: https://arxiv.org/abs/2405.11055arXiv+1ResearchGate+1

N. S. Kumar, G. Devasena, S. V. Kumar, and K. Raghu, "Voice to Text Summarization Using NLP," International Journal on Computational Modelling Applications, vol. 1, no. 2, pp. 1–9, Oct. 2024. [Online]. Available: https://www.researchgate.net/publication/390260605_VOICE_TO_TEXT_SUMMARIZATION_USING_NLPResearchGate

M. Patel and R. Kumar, "Automated Minutes of Meeting Using a Multimodal Approach," International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 13, no. 1, pp. 123–130, Jan. 2025. [Online]. Available: https://www.ijraset.com/research-paper/automated-minutes-of-meeting-using-a-multimodal-approach

Smith and B. Johnson, "IoT-Enabled Real-Time Meeting Transcription Using Deep Learning Techniques," Proceedings of the 2024 International Conference on Internet of Things and Machine Learning, Tokyo, Japan, pp. 45–52, Mar. 2024.

Lee et al., "Enhancing Meeting Summarization with Multimodal Data Fusion," Journal of Artificial Intelligence Research, vol. 68, no. 3, pp. 123–135, Jun. 2024.

Kumar and E. Chen, "Speaker Diarization and Action Item Extraction in Smart Meeting Rooms," IEEE Transactions on Audio, Speech, and Language Processing, vol. 32, no. 7, pp. 987–996, Jul. 2025.

F. Zhang and G. Li, "Real-Time Meeting Minutes Generation Using Edge Computing and NLP," Proceedings of the 2025 IEEE International Conference on Edge Computing, San Francisco, CA, USA, pp. 210–217, Apr. 2025.

H. Nguyen et al., "Automated Meeting Documentation through IoT and Deep Learning Integration," International Journal of Computer Applications, vol. 182, no. 5, pp. 25–31, May 2025.