iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57787

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Automated Minutes of Meeting Using a Multimodal Approach

Aman Singh[1], Ankit Gautam[2], Deepanshu[3], Gautam Kumar[4], Mr Lokesh kumar Meena[5], Mr Shashank Saroop[6]
Department of computer science and engineering, ADGIPS, Delhi

Abstract: ***The automated minutes of meeting using multimodal approach technique has emerged as a promising solution to lighten the time consuming and error prone manual process of capturing and summerize the meeting discussion. This research paper presents a novel approach for automating the minute of meeting through multimodal approach. The natural language processing is used to identify the different type of topic such as key topics, important discussion and other significant details discussed during meeting. The machine learning models are trained on datasets to classify and extract the relevant information accurately. Further, the research explores the use of advanced machine modals, such as whisper and transformers, to capture the context and refinement of meeting. These models enhance the accuracy and fastest of generated minutes of meeting. The assessment of the automated minutes of meeting generation involves compare of the outputs against the manually generated minutes of meeting by human notetakers. Metrics such as accuracy and F1 score are used to assess the system performance, ensuring the accuracy and quality of generated minute of meeting. This demonstrate that the automated minute of meeting using multimodal approach offers significant time savings, reduces human error, and increase overall efficiency in capturing and summarize the meeting discussion. The system shows promising potential for adoption in various organization and industry. In conclusion, this research paper present a comprehensive study about the automated generation of minute of meeting using multimodal approach. The proposed approach uses thee NLP techniques and advanced machine learning models to accurately extract and summarize meeting content. The results highlight the potential of this automated system to streamline meeting processes and enhances overall productivity in organization.***

Keywords: Natural language processing (NLP), Speech recognition, Artificial intelligence, machine learning, Whisper AI (open ai).

I. INTRODUCTION

In today business environment, meeting plays a crucial role in facilitating collaboration, decision making and information sharing among the organization. However, the process of capturing and summarization of meeting into comprehensive minutes can be time consuming and prone to human error. To address this challenge, automated generation of minute of meeting using multimodal approach has accomplished significant attention. This research paper focuses on exploring the potential of artificial intelligence. Specifically with the integration of Whisper AI, to automate the minute of meeting. Whisper AI is an advanced natural language processing (NLP) platform that clout state of the art machine learning models to analyze and understand human language. By applying Whisper AI in the context of minutes of meeting generation, we aim to devlop a system that can accurately and efficiently extract important information from meeting transcript and transform it into meaningful information. The integration of machine learning models within the Whisper AI framework enables the system to learn from annotated datasets and devlop models that can recognize important meeting elements, such as agenda items and discussion.

The research paper will delve into technical details of the implementation, outlining the various stages involved in the automated generation process. This includes cleaning and pre-processing of meeting transcripts, feature extraction, model training and evaluation.

The expected outcomes of this research include a comprehensive understanding of the capabilities and limitations of automated minute of meeting generation using multi model approach. The findings will contribute to growing body of knowledge in the field of artificial intelligence and machine learning. Additionally, the research target to demonstrate the practical application and potential benefits of this technology in various industries and organizational setting, where meeting documentation is crucial for productivity and decision making purposes.

The results of this study have the potential to revolutionize the way meeting are documented, improving overall efficiency in capturing and summarizing important meeting discussion.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

II. LITERATURE SURVEY

In order to learn in detail about this domain and the previous research done in this field, several studies and research papers were referred to. Details of some of the referred papers are mentioned below with publication time of the paper ranging from newest to oldest.

Megha Manuel et. al in 2021,mentions creating an Automated Minute Book Creation (AMBOC) model using machine learning to extract key information from important discussions in a meeting. Although the model shows good accuracy and can differentiate between male and female speakers, it does not performsatisfactorily while using other languages exceptEnglish.

Jia Jin Koay et. al in 2021,apply a sliding window approach combined with the BART summarizer to obtain summaries for evaluation. Its advantages are that it does not require annotated data and can be scaled to various domains for usage. But it shows low accuracy in some cases such as detection of highly relevant utterances.

Chenguang Zhu et. al in 2020,mention using an Hierarchical Meeting Summarization Network Model or HMNet Model which is based on an encoder- decoder transformer structure. It shows goodperformance in terms of both automated and human evaluation metrics but cannot properly cover detailed items during long meeting transcripts.

Anna Nedoluzhko et. al in 2019,analyze required tools and datasets for automatic minuting of meetings and lay out a topology of types of methods,meetings,etc.But they fail at creating a final use-able model for the purpose.

Guokan Shang et. al in 2018,propose using a framework based on Multi Sentence Compression Graph (MSCG). The method produces well framedsummaries when compared to human written transcripts but not all outputs are usable due to lack ofcoherence among several entities.

Siddhartha Banerjee et. al in 2016,propose a methodwhere supervised learning is used to separate various topic segments, and then identify the important utterances. The important utterances are then combined together. The best sub-graph is then obtained by integer linear programming (ILP) which isselected as the final output.

III. SYSTEM DESIGN

A. Speech To Text

The automated minutes of meeting using multimodal approach and Whisper AI involves a comprehensive system design that encompasses data processing and the output generation. Whisper is an automatic speech recognition (ASR) system trained on 670,000+ hours of multilingual and multitask supervised data collected from the web. Moreover, it enables in multiple language.

The Whisper architecture is a simple end-to-end approach, implemented as an encoder-decoder transformer. Input audio is split into many small chunks, converted into a log-mel spectrogram, and then passed into an encoder. A decoder is trained to predict the corresponding text caption, intermixed with special tokens that direct the single model to perform task such as language identification, phrase-level timestamp, multilingual speech transcription and to English translation.

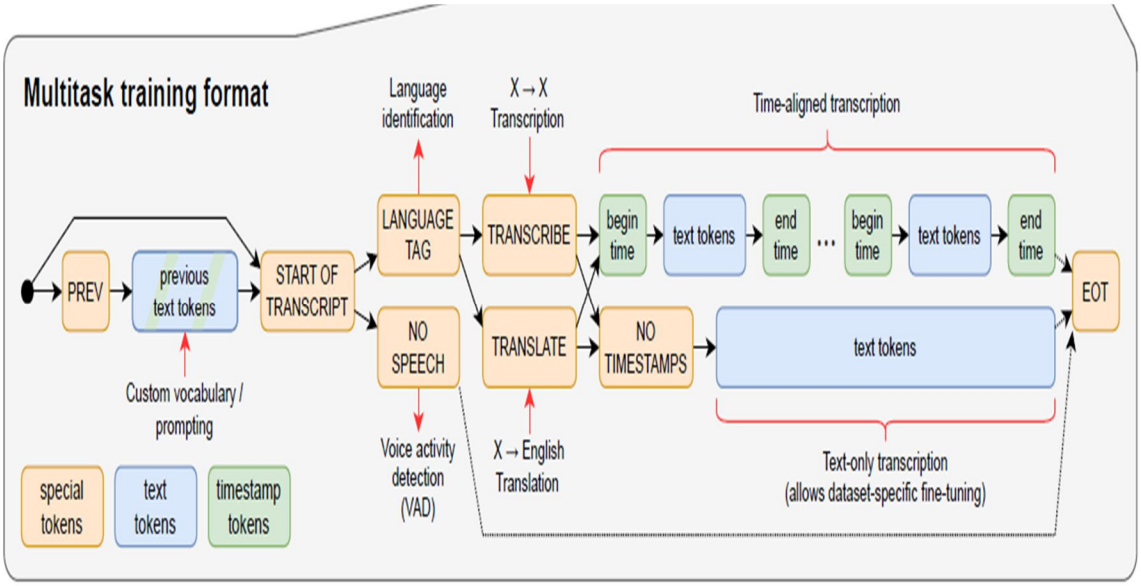


Fig. Language model to predict tex

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Speaker Identification

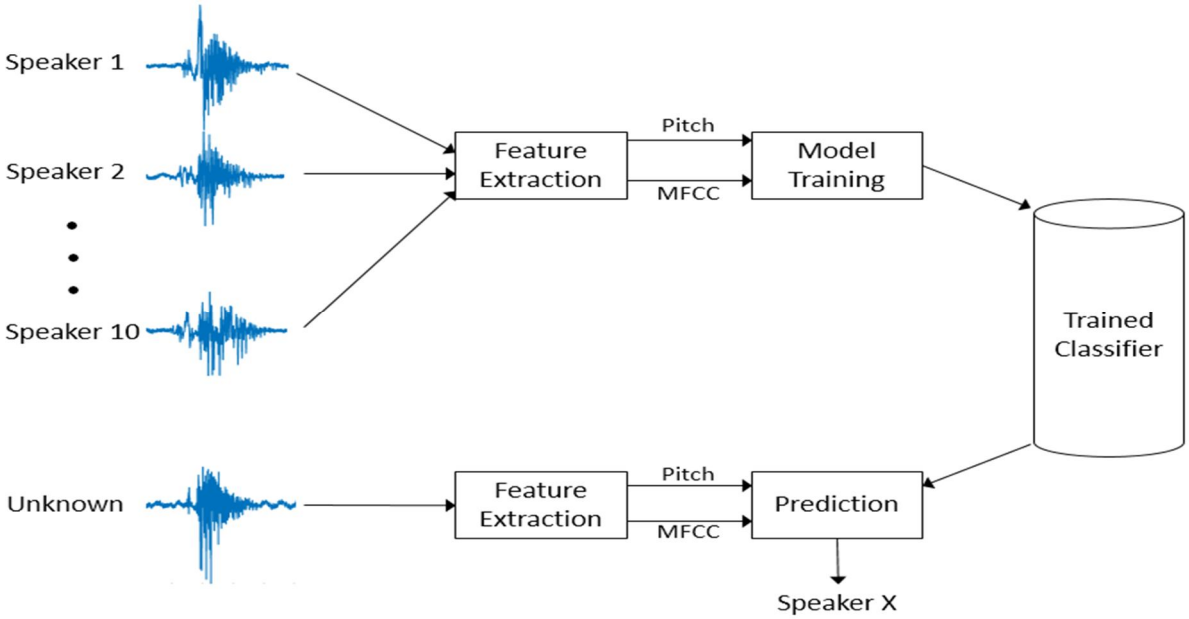


Fig. Unique audio recognition for speaker identification

MFCC, LPC, LPCC, LSF, PLP and DWT are some of the feature extraction techniques used for extracting relevant information form speech signal for the purpose speech recognition and identification. These techniques have stood the test and have widely used in speech recognition for several purposes.

IV. SYSTEM IMPLEMENTATION

Designing a speech-to-text involves several components and consideration. Here a high level overview of the system design for a basic STT system:

A. Audio input:

1) Audio file input.

2) Preprocessing: Handle audio normalization, noise reduction, and filtering.

B. Acoustic Feature Extraction:

1) Convert the audio signal into a representation suitable for speech recognition.

2) Extract features such as Mel-frequency cepstral coeffcients (MFCCs), filter, pitch and energy.

3) Apply techniques like windowing, fourier transform, and frame-based analysis to obtain time-varying features.

C. Acoustic Modelling:

1) Utilize machine learning techniques, such as deep neural networks (DNN) or convolution neural networks (CNN), to model the relationship between acoustic feature and phonetic units.

2) Train the model using a large datasets of audio records paired with their corresponding transcription.

D. Language Modelling:

1) Utilize language models to capture the statistical properties of the spoken language.

2) Incorporate linguistic knowledge, such as grammars to improve.

E. Post processing:

1) Apply post processing techniques to improve the accuracy and readability of the recognized text.

2) Handle tasks such as capitalization, punctuation restoration, homophone disambiguation, and error correction.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. RESULTS

The research paper shows detailed results and analysis of the performance of different speech recognition and summarization models. It highlights the merit and demerits of each model, identifies the factors influencing their performance, and provides insights into the trade-off between accuracy, speed, and summary details.

VI. MERITS AND DEMERITS

Clear and concise documentation of key points, discussions, and decisions made during the meeting.
Serving as a valuable reference for participants, ensuring accurate recall of information.
The automated system focuses primarily on summarization rather than capturing the extensive details and structure found in manually written minutes. it may not include specific elements such as participant names, meeting duration, and date/time in the exact format desired for official minutes.
The automated system focuses primarily on summarization rather than capturing the extensive details and structure found in manually written minutes. it may not include specific elements such as participant names, meeting duration, and date/time in the exact format desired for official minutes

VII. CONCLUSION

The research paper concludes by summarizing the brief of the comparative study and discussion the implication for future research and application in speech recognition and summarization. It highlights of selecting convenient models and techniques based on the specific use case and gives recommendations for improve the accuracy and efficiency of speech recognition. Overall, this research paper contributes to the existing body of knowledge by providing a comprehensive analysis of speech recognition and summarization models and techniques. The findings and insights can guide researchers and practitioners in selecting and developing effective solutions for speech recognition and summarization tasks.An automatic notes generator is a useful tool for creating concise, organized summaries of audio or text content. It can save time and effort by automatically identifying key points and main ideas, and organizing them into coherent notes. Automatic notes generators can be particularly useful for students and professionals who need to take notes from lectures, meetings, or other spoken or written materials. They can also be useful for language learners, as they can help to extract the most important information from listening or reading materials.There are a number of different techniques and approaches that can be used to build an automatic notes generator, including speech recognition, natural language processing, and text summarization. The specific method or combination of methods used will depend on the specific requirements and goals of the application. Notes is an important part of daily meeting, classes and seminars, it is a must to write notes during attending the meeting and keeping track of all points discussed and express new ideas and suggestion in meeting. But sometimes it is difficult to write notes simultaneously listening to points and the main concept is to keep the points, agenda of meeting to be kept safely. There are indeed several challenges and limitations to speech-to-textconversion technology, which is a form of natural language processing that involves transcribing spoken language into written text.

.

VIII. FUTURE SCOPE

Whisper also supports transcription of various other languages, like Hindi, Marathi, Tamil,etc and auto translation. We can use this feature to create a system that can generate minutes even from bilingual meetings making it ideal for a country like India where most of the people speak more than one language.
Currently we use a transformer which has an input token limit of 1024, out of which the first and last tokens are special tokens so effectively we can only process 1022 tokens at a time. However, in future with the recent developments in longformers, we can greatly increase the amount of tokens that can be fed to model at a time. This will greatly increase the quality of summaries generated.
A proper minutes of meeting usually has a proper format. Like the name of participants, duration and date/time of meeting so we can use template filling to create proper minutes of meetings. We can combine our system with a powerful Named Entity Recognition model to extract names of participants to create minutes of meeting with proper format as one would create by hand.
In future we can deploy our model so that everyone can take its advantages and save time and effort of creating minutes by hand.

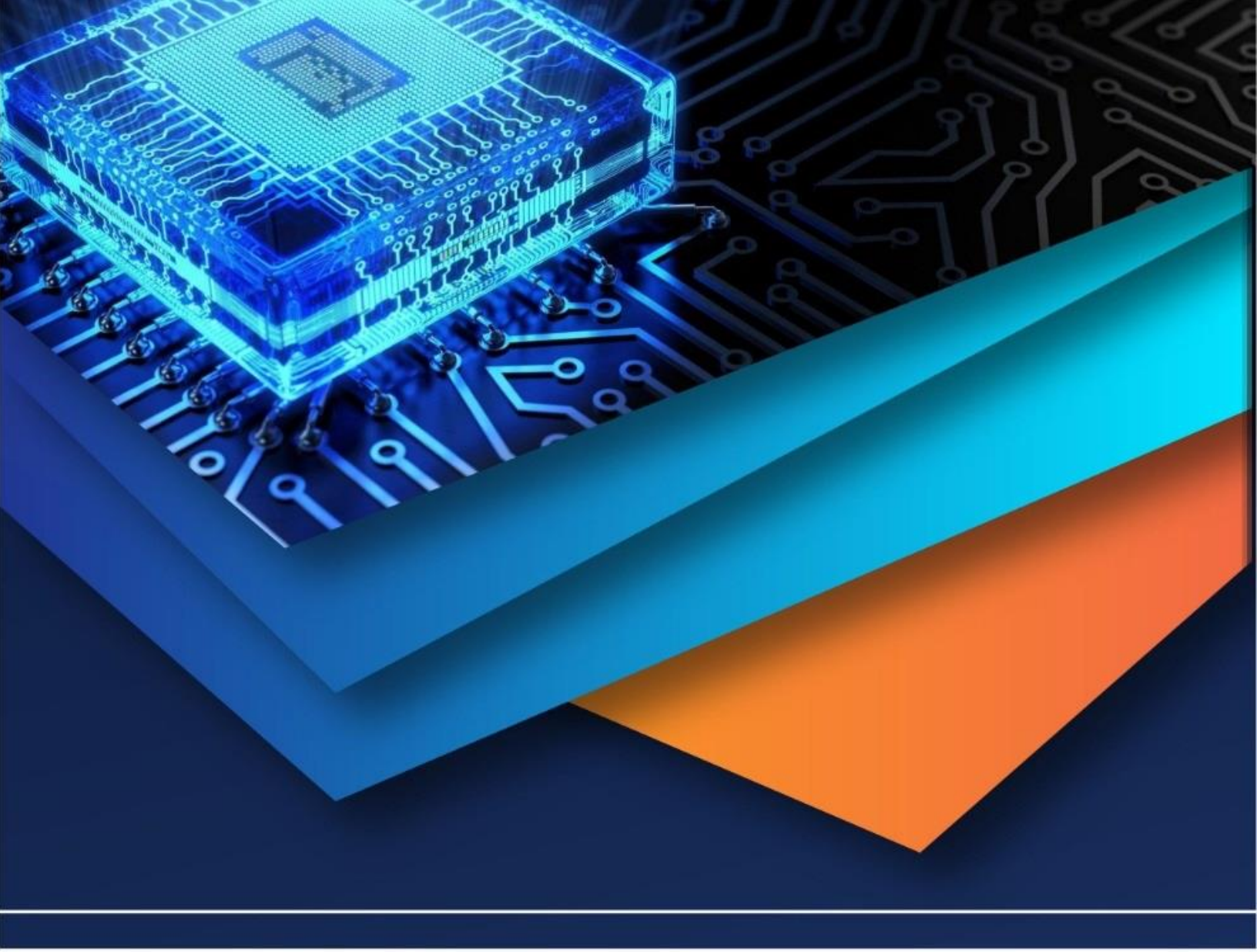
REFERENCES

[1] A.M. Attorney. The Corporate Records Handbook: Meetings, Minutes and Resolutions. dElta printing solutions, inc., 2007.
[2] A.M. Attorney. Nonprofit meetings, minutes, and records. Delta printing solutions, inc., 2008.

©IJRASET: All Rights are Reserved

[3] S. Banerjee, P. Mitra, and K. Sugiyama. Abstractive meeting summarization using dependency graph fusion. In Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion, pages 5–6, New York, NY, USA, 2015. ACM.
[4] S. Banerjee, P. Mitra, and K. Sugiyama. Generating abstractive summaries from meeting transcripts. In Proceedings of the 2015 ACM Symposium on Document Engineering, DocEng '15, pages 51–60, New York, NY, USA, 2015. ACM.
[5] S. Banerjee, C. Rose, and A.I. Rudnicky. The necessity of a meeting recording and playback system, and the benefit of topic–level annotations to meeting browsing. In M.F. 11http://elitr.EU/ Costabile and F. Paternò, editors, Human-Computer Interaction - INTERACT 2005, pages 643–656, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.
[6] F. Bargiela and S.J. Harris. Managing Language: The discourse of corporate meetings. John Benjamins Publishing Company, 1997.
[7] F. Boudin and E. Morin. Keyphrase extraction for n-best reranking in multi-sentence compression. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 298–305, Atlanta, Georgia, June 2013. Association for Computational Linguistics.
[8] T. Bui, M. Frampton, J. Dowding, and S. Peters. Extracting decisions from multi-party dialogue using directed graphical models and semantic similarity. In Proceedings of the SIGDIAL 2009 Conference, pages 235–243. Association for Computational Linguistics, 2009.
[9] Susanne Burger, Victoria MacLaren, and Hua Yu. The isl meeting corpus: The impact of meeting type on speech style. 01 2002.
[10] S.E. Butler. Mission Critical Meetings: 81 Practical Facilitation Techniques. Wheatmark, 2014.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)