International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Conversational Analytics System

Kilambi Pavanadithya[1], Malliboyini Venkat Kalyan[2], Yatakalla Sarika[3] and Gondra Jahnavi[4] ,Mr. Jagadeeshwar Reddy V[5]

[1, 2, 3, 4] *Department of CSE(Data Science), AVN Institute of Engineering and Technology* under the guidance of
[5] *Asst. Professor, Department of CSE(Data Science), AVN Institute of Engineering and Technology*

Abstract-*With the rapid rise in spoken interactions happening in digital spaces, organizations, and research settings, there's a growing demand for reliable and automated conversation analytics. The goal of conversation analytics is to turn unstructured speech data into meaningful, structured insights, which helps us gain a better understanding of dialogues involving multiple speakers and the content being discussed. Thanks to recent advances in speech and language technologies, we're now able to analyze conversational audio more accurately and with better context than ever before.*

This paper focuses on how conversation analytics can transform raw speech into valuable analytical information using techniques like automatic speech recognition, speaker diarization, and natural language processing. These methods help identify who's speaking, transcribe what they're saying, and pull out linguistic and contextual patterns from conversations. The insights we get from analyzing this conversational data enhance our understanding of communication dynamics and support better decision-making in various fields. Overall, this study underscores the increasing academic and practical importance of automated conversation analysis, especially when it comes to managing large volumes of speech data effectively.

***Index Terms*—Automatic Speech Recognition, Conversation Analytics, Natural Language Processing, Speaker Diarization, Speech Processing**

I. INTRODUCTION

One of the easiest ways we connect with each other is through talking. Conversations are super important in settings like meetings, customer service, education, healthcare, and media. With more people chatting on digital platforms, tons of conversational audio is generated every day. But going through all that audio by hand is tough and takes a lot of time, especially with long discussions featuring different speakers. Because of this, automated conversation analytics has become a key area of research.

Conversation analytics is all about pulling valuable insights from spoken interactions using techniques like speech processing and natural language processing. Unlike traditional audio analysis, which mainly focuses on the signals, conversation analytics digs into the flow of ideas, who's talking, and the sequence of what's being said. Some of the main parts of conversation analytics include automatic speech recognition, figuring out who's speaking, and linguistic analysis. By blending these elements together, we can create a structured model of conversational data.

Lately, breakthroughs in deep learning have really boosted how well speech recognition and speaker identification systems work. Now, we can understand real-life conversations even when people talk over each other, when there's background noise, or when speakers have different accents. Plus, advances in natural language processing have allowed us to dive deeper into the spoken words—for example, spotting sentiment, figuring out intentions, and pinpointing context. These upgrades have made automated conversation analysis much more reliable and scalable, which is why it's gaining popularity in both research and practical use.

But conversation analytics is more than just turning speech into text. It gives us insights into communication styles, personality traits of speakers, and overall interaction patterns. These insights are really helpful for figuring out how effective conversations are, understanding group discussions, and helping with decision-making. As the need for smart speech analysis tools increases, conversation analytics is becoming an essential field that transforms raw audio into meaningful insights.

II. LITERATURE SURVEY

Over the past two decades, there's been a lot of exciting progress in the field of speech and language processing, paving the way for the conversation analytics systems we have today. Early contributors like Jurafsky and Martin[1] really set the

groundwork, establishing a solid theoretical base for things like speech recognition and natural language processing, as well as discourse analysis and spoken dialogue systems. They identified some key principles to grasp conversational speech, including aspects like acoustic modeling, language modeling, and how conversations are structured. These principles are all critical for automatically analyzing spoken interactions.

A big focus in conversation analysis has been on Automatic Speech Recognition (ASR). The arrival of deep neural networks for acoustic modeling has truly transformed the field, leading to impressive improvements in ASR performance. Hinton et al.[7] showed that deep learning models could significantly outperform older techniques, resulting in better accuracy and reliability. Building on this, Radford et al.[2] created a large-scale weakly supervised speech recognition model that can work well with different accents, noisy settings, and multiple languages. This has made ASR much more effective in real-life conversations, allowing for accurate transcriptions of long and complicated audio discussions.

Speaker diarization, which aims to figure out who spoke when in a conversation. Early studies by Anguera et al.[8] looked at traditional approaches and pinpointed issues like overlapping voices and variability in speech. Lately, deep learning advancements have taken diarization to the next level. For instance, Bredin et al.[4] introduced a neural framework that acts as a versatile toolkit for speaker diarization, enhancing our ability to segment and identify speakers in group chats. These improvements have really bolstered the reliability of diarization in real-world audio data.

Natural Language Processing (NLP) plays a crucial role in pulling out meaningful insights from transcribed speech. Liu[9] laid down foundational techniques for sentiment analysis and opinion mining, which remain vital for interpreting how speakers feel and the emotional undertones in text. Modern NLP frameworks, like the efficient toolkit developed by Hannibal et al.[3] , have streamlined large-scale language analysis, offering solid processes for tasks like tokenization, syntactic parsing, and text classification. These resources make it easier to analyze conversation transcripts on a larger scale.

On the emotional side of things, there's growing attention on analyzing affective and conversational context. Poria et al.[5] have delved into affective computing, stressing the importance of studying emotions and contextual signals in human communication. They pointed out that grasping sentiment and emotional trends can provide deeper insights into how conversations unfold. Furthermore, Stocked and Shriberg[10] looked into statistical methods for spoken dialogue systems, focusing on how to model the structure of conversations, who speaks when, and how interactions flow. These aspects are crucial for understanding conversations more deeply.

Audio preprocessing and feature extraction are the starting points for any speech-based system. McFee et al.[6] came up with a detailed Python framework for analyzing audio signals, giving the essential tools needed to manage and process raw audio before diving into more complex speech and language analysis. These preprocessing techniques lay the groundwork for reliable follow-up tasks like diarization and speech recognition.

There’s been a lot of progress in areas like speech recognition, speaker diarization, and text analytics, many researchers have been looking at these components in isolation. The work is ongoing in trying to bring these techniques together into a unified conversation analytics framework. This highlights the need for systems that can merge audio processing, speech recognition, speaker identification, and linguistic analysis to truly understand conversational data.

III. PROPOSED SYSTEM

The proposed system introduces a Conversation Analytics System that can effectively automate the processing and understanding of spoken conversations by means of speech and language processing techniques. The web-based application allows users to authenticate themselves and thereby be able to analyze conversational audio by either uploading recorded files or performing live voice recording. Thus, the system intends to transform unstructured audio data into analytical insights that better facilitate understanding of the conversational content.

In the activity flow diagram, it is shown that the system initiates from user interaction via a web interface. Users go through the process of registration or logging in to the system in a secure way and thus get access to the main dashboard. From the dashboard, users decide on the input mode—uploading a file or using live recording. The audio selected is sent to the controller module that handles the entire processing pipeline.

The system first conducts audio preprocessing to ready the raw audio for the next stages of the analysis. Noise reduction and segmentation are some of the steps taken during preprocessing to make sure that the processing remains reliable. After preprocessing, relevant features of the audio are extracted and then sent to the speech analysis pipeline along with the processed audio. To convert vocal utterances into textual form, automatic speech recognition is implemented, making it possible for the linguistics analysis to continue.

For multi-speaker conversations, the system has speaker identification and diarization capabilities that enable it to detect and separate individual speakers when the conversation has taken place. This means that it will be perfectly clear who said what. Along the same lines, the text transcript is subjected to natural language processing wherein its sentiment and other emotional components can be analyzed, thereby understanding the contextual tone of the conversation.

The different components involved in the processing are housed in one place, a voice analysis engine, as shown in the figure of the system architecture. The analytical results such as speaker-wise transcription, sentiment markers, and conversational summaries are collated into one analytics report. The conversation history is archived, and future access is facilitated through the storage of the original results in a database.

In the end, the insights obtained from the analysis are brought to the user by means of an interactive dashboard that makes it possible to view and interpret the conversation analysis in a clear way. Moreover, the system offers a feature that lets users export or download the generated report. The sequence diagram indicates the interaction among the web interface, authentication service, controller, processing modules, and analytics component that is performed in an orderly, secure, and efficient manner.

The system is leveraging audio processing, automatic speech recognition, speaker identification, and text analysis which are all combined in a modular and scalable framework for an integrated and fully automated conversation analytics solution. The system acts as a good link between unprocessed conversational audio and well-structured analytical insights.

IV. TOOLS and TECHNOLOGIES

The Conversation Analytics System will be written mainly in Python. The Flask web framework will be utilized for the system backend, which will be responsible for user interaction and processing workflows management. The front-end interface will be a product of HTML, CSS, and JavaScript that will be capable of supporting audio uploading, live recording, and providing visualization of the analytics results.

The audio preprocessing and signal handling tasks will be done through the use of Python-based audio processing libraries. Neural-based diarization methods will be used for speaker diarization, and deep learning–based speech recognition models will be used for automatic speech recognition to convert speech into the text. Natural Language Processing methods will be employed to analyze the copied text to get the conversational insights. A lightweight database system will serve as a place to keep user information, conversation data, and analysis results.

V. FIGURES

Figure 1- System Architecture

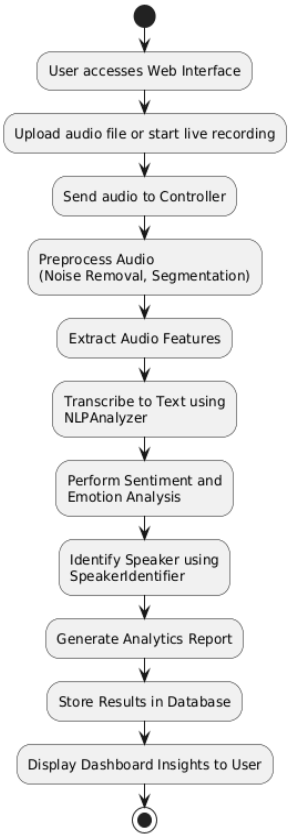

Figure 2- Activity Flow Diagram

VI. CONCLUSION and FUTURE WORK

Here, a Conversation Analytics System has been introduced through this paper. It centers on automated analysis of spoken conversations. The aim is to extract meaningful and structured insights from audio data. The proposed system is essentially a conversation analytics tool that deals with unstructured speech, i.e., it applies speech recognition, speaker identification, and language analysis techniques to a human speech. By converting conversational data into easily understandable analytical results, the system facilitates better comprehension of communication patterns and conversational dynamics. The paper effectively conveys the significance of automated conversational analytics within speech and language processing research and lays down a solid theoretical framework for the analysis of real-world conversational data.

The next stage of work on the introduced system will be to continue its development and convert it into a fully-fledged web-based application by means of Flask and other supportive Python technologies. Future upgrades will therefore be specifically directed towards consolidating the tying of the system's components and enhancing the analytical functionality of the speech and language processing modules. The system's effectiveness and user-friendliness will be improved by including more conversational features and polished analysis methods.

VII. REFERENCES

[1] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Pearson Education , 2023.

[2] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust speech recognition via large-scale weak supervision,” Proceedings of the International Conference on Machine Learning (ICML), 2023.

[3] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spaCy: Industrial-strength natural language processing in Python," Zenodo, 2020.

[4] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, et al., "pyannote.audio: Neural building blocks for speaker diarization," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020.

[5] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," Information Fusion, vol. 37, pp. 98–125, 2017.

[6] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in Python," Proceedings of the 14th Python in Science Conference, 2015.

[7] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, et al., "Deep neural networks for acoustic modeling in speech recognition," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.

[8] X. Anguera, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, and O. Vinyals, "Speaker diarization: A review of recent research," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 2, pp. 356–370, 2012.

[9] B. Liu, Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

[10] A. Stolcke and E. Shriberg, "Statistical approaches to spoken dialogue systems," in Handbook of Speech Processing. Springer, 2006