ISSN No:-2456-2165

# Meeting Insights Summarisation Using Speech Recognition

Sakshil Verma[1]
Computer Science and Engineering Student,
SRM Institute of Science and Technology
Chennai, Tamil Nadu, India

Saksham Thareja[2]
Computer Science and Engineering Student,
SRM Institute of Science and Technology
Chennai, Tamil Nadu, India

Dr. P. Supraja[3]
Associate Professor, Department of Networking and Communications, SRM Institute of Science and Technology,
Chennai, Tamil Nadu, India

**Abstract:- Speech is the strongest mode of discourse through which people express their emotions and ideas through numerous languages. Speech recognition authorization has varied applications as it provides Hassle free procedure which does not require physical contact as in the case of fingerprint authorization. Speech summarisation methods use speech from people as input and produce a condensed form as spoken or written language. Speech synthesis offers a variety of applications spanning from computer technology to medical care, including improving language libraries and reducing therapeutic paperwork load. Every dialect has its unique collection of features for speaking. Despite speaking a comparable language, the speed and dialect differ from individual to individual. This can make comprehending the conveyed message difficult for certain people. Conferences are an important part of every organisation's operation, regardless of if they took place via the web or in reality. Meeting translation and summarization standards, on the contrary hand, are typically disagreeable demands because they necessitate time-consuming workers. This project aims to identify things during meetings like the greatest number of times a person spoke in a meeting to determine his level of inputs and summarisation of insights of meetings for all the employees in the meeting and identifying their insights through the words spoken by them.**

***Keywords:-*** *Speech Recognition, Speech Summarization, Speech Pre-Processing, Spacy, Gensim.*

## I. INTRODUCTION

Speech is a highly powerful mode of communication through which humans express their thoughts and feelings through numerous languages. Each language has its unique set of linguistic qualities. Even while speaking the same language, the speed and accent vary from person to person. It makes it difficult for certain people to comprehend the conveyed message. Long speeches can be difficult to follow at times owing to factors such as differing pronunciation, pace, and other factors. Speech recognition, which is a cross-disciplinary issue in computational language science, contributes to the advancement of technology that allows for the recognition and translation of speech into text. Text summarization pulls the most significant information from a text-based source and offers an effective summary of the same.

Speech summarization is the process of condensing human speech into a more concise and manageable form. It tries to write a summary that is suitable for a specific task. The summary should be more coherent than a direct transcription of speech, as it eliminates common irregularities, breaks, repairs, and repetitions. The recent interest in speech summarization is driven by improvements in improving the precision of speech recognition systems, the standard in audio capturing, and the rising use of natural language as a computer structure.

The process of speech summarization involves several technological components such as automated speech recognition (ASR), which translates voice into written form, and summary modules, which summarise information summarise key parts of the transcription. Users can use the Internet's Voice APIs to capture audio and submit it to a speech recognition web service for processing.

Speech summarization has a range of real-world applications, such as summarising broadcast news, podcasts, clinical conversations, and meetings. It presents a challenge in speech understanding research and can be achieved through extractive or abstractive summarization techniques. Extractive summarization preserves the original format and is typically more fluent, while abstractive summarization is more concise and flexible. The summary of speech ought to be more intelligible than a straight transcript.

Meetings are a common and important part of business operations. They provide opportunities for team members to collaborate, exchange ideas, and make decisions. However, meetings can also be time-consuming and distracting, making it difficult for attendees to retain key information and insights. To address this challenge, the use of speech recognition and summarization technology has gained attention as a way to efficiently and effectively process meeting content.

ISSN No:-2456-2165

Voice recognition is a technique that converts words that are spoken into text. It is also known as automated speech recognition (ASR) or speech-to-text. It is an interdisciplinary field that involves speech signal processing, acoustic modelling, language modelling, and machine learning. The goal of speech recognition is to accurately transcribe. human speech into written text, allowing for easier processing, storage, and retrieval of spoken information.

Recognition of speech has an extensive variety of use cases, including voice-activated artificial intelligence, dictation software, and voice recognition software, accessibility solutions for people with disabilities, and hands-free control of devices.

Speech recognition is a challenging task due to the complexity of human speech and the variability of spoken language. Some of the major challenges include:

- *Speaker Variability*

Different speakers have unique speech patterns, including pronunciation, speaking rate, and intonation. This variability can make it difficult for speech recognition systems to accurately transcribe speech from different speakers.

- *Background Noise*

The existence of ambient noise can decrease the quality of the spoken signal dramatically and make it more difficult for the system to accurately transcribe the speech.

- *Vocabulary Size*

The size of the vocabulary that a speech recognition system needs to support can have a significant impact on its accuracy. Larger vocabularies require more complex language models, which can be more difficult to train and can result in lower recognition accuracy.

Speech pre-processing consists of reducing background noise, adjusting loudness, and transforming the speech input to a digital representation. The process of extracting features from a Speech signal entail translating it into a collection of distinguishing qualities that may be used to identify the words uttered.

Speech summarization approaches are often based on a mix of the processing of natural language (NLP), recognition of speech, and machine learning are all examples of artificial intelligence (AI). The precision of detection of speech engine, the quality of the NLP algorithms, and the efficacy of the machine learning models utilised all influence the quality of the summary output. With the continuous progress in Speech recognition accuracy and the rising appeal of natural language for a computer gateway, there has recently been a spike of interest in speech summarising approaches.

To summarise, Speech summarization is a difficult process that necessitates the use of a mix of speech recognition, NLP, and machine learning approaches. Speech summarization's purpose is to give a shortened and more intelligible version of the speech that is appropriate for a certain activity. The two primary types of speech summary approaches are extractive summarization and abstractive summarization, each with its own set of advantages and disadvantages. The accuracy of the Speech recognition engine, the quality of the NLP algorithms, and the efficacy of the machine learning models utilised all influence the quality of the summary output.

## II. LITERATURE SURVEY

Converting speech to text is beneficial in a variety of scenarios. Jose et al. developed an effective technique for obtaining English fluency that improves the user's speaking style through proper pronunciation using English phonetics. Sivakumar et al. did a comparative study of the advantages and disadvantages of different sizes of vocabulary Voice recognition systems. The research conducted highlighted the significance of computational models of language in enhancing the precision of monologue-to-text translation across various interference and breached-word conditions. Yogita and co-workers developed a bilingual language conversion technology using the extraction of features from MFCC and audio classification algorithms such as the Least Length Encoder and Support Vector Machine (SVM). Sphinx 4, a platform that is free to use, was recommended for converting authentic Bengali text into English. In the information set beneath examination, the researchers estimate to have achieved a level of precision of 71.7%. Wan proposes summarising English text using association semantic criteria. The novel extraction approach, according to the author, shows enhanced extraction convergence and precision. LDA is the most extensively used topic-based text categorization algorithm.

A novel method to similarity calculations suggests a change for the better. Saiyed and Sajja gave a succinct summary of the various categories of summarising methodologies, emphasising their advantages and disadvantages. This work offers researchers advice on selecting particular methods in accordance with their requirements. Choosing the right term is a multi-objective optimization problem. With this, the writers applied a human-centred training optimisation technique. According to the authors of, feature extraction using neural networks is more effective than online extractive techniques. Vythelingum et al. proposed a method for detecting errors in grapheme-to-phoneme conversion in speech-to-text generation. Authors stated that the method they used had a greater rate of mistake adjustment, therefore would help the real-life annotator. As stated in the scientific review that resulted in this study's activity, the transformation of voice to written form and its summation are essential. A cross-dimensional text summarising technique based on dimensional selection and filtering was proposed by Zenkert et al. Using the findings from the Multidimensional knowledge representation database, the technique was evaluated. Devasena and Hemalatha's content processor was utilised to identify the arrangement of the content that was entered.[1] Transcribing spoken word materials including

ISSN No:-2456-2165

speeches, presentations, lectures, and news broadcasts is one of the main uses for automatic voice recognition [2].

Although speech is the most efficient and natural form of human communication, just recording speech as an audio signal makes it difficult to quickly examine, retrieve, and reuse speech documents. Speech transcription is therefore anticipated to be a key skill in the upcoming IT era. Notwithstanding the reality that extremely high identification accuracy can be readily achievable for voice given via a written work, such as anchoring commentators' news conference phrases, the ability of tech to distinguish speech that is impulsive remains limited. [3]. Only one survey work evaluating various output summaries, features, methodologies, and assessment criteria has been published on automatic speech summarization [4]. The present research focused solely on a two-phase summarising approach that included essential phrase retrieval and phrase compression, and it mainly evaluated at publications released around the year 2006. In the year 2008, the exact same researchers published another investigation of unstructured speech recordings that addressed issues such as audio collections, pronunciation recognition, auditory simulations, language structure, the process of extraction, and voice synthesis [5].

The bulk of the initial research on separate-document summarization was focused on scientific papers. The most widely cited paper on synthesis is likely the first (Luhn, 1958), and that discusses studies undertaken at IBM in the early 1950s. According to Luhn's study, the number of times of a certain phrase in a piece of writing is a fair measure of its significance.

Some major ideas advanced in this research have gained prominence in subsequent work on summarization. Terms were initially rooted to their fundamental kinds, and then the endings were deleted. Luhn then created a list of keywords and phrases arranged by frequency that were reduced, with the ranking supplying an indication of the phrase’s relevance. On an expression level, an importance component was established that reveals the total number of repetitions of noteworthy words inside an expression, in addition to the standard deviation separating them due to not important word interventions. Each of the phrases are scored relative to their significance element, and those with the highest scoring statements are subsequently selected to construct the activate-abstract. A comparable study (Baxendale, 1958), additionally conducted at IBM and presented in the very same journal, gives an early glimpse into a key attribute beneficial in spotting major portions of papers, notably phrase placement. This writer investigated 200 segments to reach this goal and determined that the topic phrase occurred as the initial phrase in 85% of the sentences and as the final word in 7% of the subsections. As a result, identifying one of each of these is a basic but somewhat precise approach of identifying the subject phrase. This geographical feature is now employed in a number of complicated artificial intelligence applications. [6]

A summary, according to Radev et al., is "a text that is formed from one or more texts, that conveys\important information in the original text(s), and that is no\longer than half of the original text(s) and usually, \significantly less than that". The summarization of a text is the process of identifying and seeking the key and most notable details within a piece of writing or set of related writings, and subsequently distilling it into a simpler form that maintains the basic idea. The act of creating a short and flowing synopsis that preserves the essential as well as general significance is known as summarization of text automatically. [7]

In 2015, Nallapati et al. used methods involving deep learning in abstracting and summarising texts for the very first occasion, and the suggested methodology centred on the encoding and decoding framework.

The encoder-decoder models were designed to solve Sequence to Sequence difficulties (Seq2Seq). The initial pattern of the artificial brain is translated into a comparable pattern of characters, phrases, or sentences using Seq2Seq algorithms. This approach is employed in many NLP uses such as machine interpretation and summarization of content. The list of inputs in the content condensing is the data that needs to be summarised, and the order of results is the summary that is produced. [8]

The following is the hypothesis suggested by X. Wan et al.: The first step in the reverse parser generates an explanation spanning right to left, similar to the Seq2Seq-Attn model. 2. Both the encoding device and the reversing processor employ the focus approach so that the forward-looking processor may construct an overview from left to right. Both the forward as well as backwards decoding algorithms utilise a pointer-based approach.[9]

## III. SPEECH RECOGNITION

Voice recognition is a technique that is sometimes known as automated recognition of speech (ASR), used to convert spoken words into written or transcribed text. The technology has made significant advancements in recent years, driven by improvements in machine learning algorithms, speech recognition accuracy, and audio capture quality. The Web Speech API is one of the latest developments in speech recognition technology and enables people to capture mic sounds and submit it via a speech detection web page for analysis. The API provides developers with the ability to add speech-to-text functionality to their applications.

Voice recognition is utilised in numerous applications particularly operated by voice systems, personal assistants, hands-free dictation systems, and call centre automation. The accuracy of speech recognition systems has considerably improved over the past decade as a result of breakthroughs in computer learning and neural networks with deep layers.

ISSN No:-2456-2165

The field of speech recognition has advanced significantly in the past few years, and it is now possible to transcribe speech with high accuracy, even in noisy or reverberant environments. This makes it possible to use speech recognition technology to facilitate the summarization of office meetings, which can save time and effort, and allow for the quick and easy dissemination of key information from the meeting.

*A. Components of Speech Recognition Systems:*

Speech recognition systems are composed of several components that work together to transcribe speech into text. STT, often known as Voice recognition, is a method of converting spoken words into printed text. The purpose of STT is to translate spoken words as correctly and fast as feasible into machine-readable format.

Speech pre-processing, feature extraction, acoustic modelling, language modelling, and decoding are all components of the STT process:

- *Speech Pre-Processing*:

The raw audio signal is processed beforehand to eliminate undesirable noise and distortions and to improve its quality for better speech recognition performance.

- *Feature Extraction:*

This component processes the raw speech signal to extract relevant information that is used to identify the words spoken. This includes processing to remove noise, normalise the signal, and extract features such as spectral coefficients, prosodic features, and pitch.

- *Acoustic Modelling:*

Acoustic modelling involves training machine learning algorithms on large amounts of speech data to recognize patterns in the speech signal and identify the sounds that make up speech. The resulting model is then used to transcribe new speech. This component uses the features extracted from the speech signal to model the sound patterns of spoken words. This typically involves training machine learning algorithms, such as Hidden Markov Models (HMMs) or Deep Neural Networks (DNNs), on large amounts of speech data to learn the relationships between the acoustic features and the spoken words. The extracted features are used to train an acoustic model, which maps the features to a set of possible phonemes or sub-word units.

- *Language Modelling:*

A language model is used to model the relationships between the acoustic models and the words in a language. It is used to predict the most likely word sequences given the acoustic models. Language modelling involves considering the context of the words being spoken to increase the precision of the STT system. For instance, the STT system may be capable of recognizing that a word is more likely to be "bank" as a financial institution rather than a riverbank, based on the words that have been spoken before. This component uses statistical techniques to model the structure of a language, including the probabilities of word sequences, grammar, and pronunciation. This information is used to disambiguate between words with similar pronunciations, and to choose the most likely transcription given the speech input.

- *Decoding:*

Decoding involves using the acoustic and language models to transcribe the speech signal into written text. The acoustic models and the language model are combined to generate the final recognized text. The decoder outputs the most likely word sequence based on the acoustic and language models like a hypothesis, which is the most likely transcription of the speech.

*B. Speech Reconnaissance System Types:*

Speech-Based Recognition Systems are classified into two distinct categories:

- *Isolated Word Recognition:*

This type of system is designed to recognize a limited vocabulary of isolated words, such as "yes" or "no". It is often used in applications such as Speech-activated controls, where the user is required to speak a limited set of predefined words.

- *Continuous Speech Recognition:*

This type of system is designed to transcribe speech in real-time, without requiring the user to pause between words. It is used in applications such as dictation software and Speech-activated virtual assistants, where the user is expected to speak naturally and continuously.

STT Technology has advanced significantly in recent years and continues to do so and is used in many applications, such as voice-activated virtual assistants, voice-activated TV remotes, voice-controlled devices, call centres, and speech-enabled accessibility technologies.

However, despite the advances in technology, STT systems can still be inaccurate, especially when dealing with different accents, noisy environments, or fast speech. The size of the vocabulary that a speech recognition system needs to support can have a significant impact on its accuracy. Larger vocabularies require more complex language models, which can be more difficult to train and can result in lower recognition accuracy. The ongoing research and development in this field aim to improve the accuracy and speed of STT systems, making speech recognition an increasingly important technology for the future. [10]

## IV. SPEECH SUMMARISATION

- *Overview:*

Speech summarization is the process of reducing the length of a speech while retaining its most important content. Summarization techniques are methods used to condense text into a more manageable form. The goal of speech summarization is to provide a condensed and more understandable version of the speech that is suitable for a specific task.

- *Speech Summarising Approaches are Classified Into two Types:*

➢ *Integrative Summarising:*

Choosing important words is an example of summarised extracts, sentences from the original text to create a summary. This approach includes picking and retrieving some of the most significant phrases from an article. Extractive summarization preserves the format of the original speech and is usually more fluent but can result in a summary that is less concise. A summary is formed by combining chosen sentences, that retains the main points and key information from the speech. Extractive summarization is mainly used for tasks where preserving the original format is important, such as legal documentation, news articles, etc.

➢ *Abstractive Summarization:*

This approach includes creating fresh phrases that summarise the main points of the speech. The new sentences are created by using a combo of Artificial learning and the processing of natural languages (NLP) approaches. Abstractive summarization is more concise and flexible, but it is also more complex and harder to implement than extractive summarization. Abstractive summarization is mainly used for tasks where summarising the speech in a more concise manner is important, such as generating executive summaries, summarising long conversations, etc.

➢ *Meeting Insights Summarization:*

The proposed solution for meeting insights summarization involves the use of speech recognition and summarization techniques. First, speech is recorded and transcribed into text using ASR. Next, summarization techniques are applied to the transcribed text to condense the information into a more manageable form. The goal of this approach is to provide attendees with a summary of the meeting's key information and insights, allowing them to more effectively retain and recall the content of the meeting.

This technology can be used to facilitate the summarization of office meetings by automatically transcribing the speech into text, which can then be processed by a summarization algorithm.

- *The Process of Office Meeting Summarization Using Speech Recognition can be Broken Down into the Procedures that follow:*

✓ *Speech Recognition:*

The initial step is to type the speech from the office meeting into text. This can be done using speech recognition software that converts the audio of the speech into a text representation.

✓ *Text Processing:*

The transcribed text is then processed to remove any redundant or irrelevant information, such as filler words, repetitions, or irrelevant comments.

✓ *Keyword Extraction:*

The processed text is then analysed to extract the most important keywords and phrases that capture the essence of the speech.

✓ *Summarization:*

The extracted keywords and phrases are then used to generate a concise and coherent summary of the office meeting. This summary can be in the form of a written document, or a presentation, or a summary report.

✓ *Review and Refinement:*

Finally, the generated summary is reviewed and refined to ensure that it accurately reflects the content of the office meeting and that it is clear and concise.

➢ *Gensim*:

Gensim serves as a freely available processing of natural languages and a subject modelling framework. One of its core functionalities is text summarization. Gensim's summarization module provides an implementation of the TextRank algorithm, which is a graph-based approach to extractive text summarization. [11]

The TextRank algorithm starts by splitting the input text into sentences and constructing a graph where the vertices represent sentences and edges show the resemblance among them. The resemblance of phrases is often calculated using word coincide, co-occurring or cosine correspondence. The TextRank algorithm is applied when the graph has been built, applying PageRank, a well-known algorithm for finding the importance of nodes in a graph, to the vertices (sentences) in the graph. The result is a ranking of the sentences, with the most important sentences having the highest score.

Finally, the Gensim summarization module selects the top-k sentences with the highest scores, where k is a user-defined parameter, to form a summary. The resulting summary gathers the most important data from the supplied text, while omitting redundant or irrelevant information.

As a supplement to the TextRank algorithm, Gensim supports alternative synthesis approaches such as Non-negative Matrix Factorization, Latent Dirichlet Allocation and Latent Semantic Analysis. These techniques can be used to generate summaries based on the underlying topics and latent structures in the text.

In conclusion, Gensim summarization is a powerful tool for generating concise and meaningful summaries of large amounts of text. Its TextRank algorithm design delivers a simple yet efficient way for extracting the most significant data off the text being entered.

➢ *Spacy:*

Spacy serves as a freely available Python toolkit for sophisticated natural language processing. It is intended to be quick, efficient, and simple to use. Part-of-speech tagging, tokenization, dependency parsing, named entity

ISSN No:-2456-2165

identification, text categorization, and other text analysis and manipulation functions are available in Spacy. [12]

Spacy's quickness constitutes one of its primary assets. It is designed for massive-scale processing of data and can swiftly and effectively handle enormous quantities of information. That makes it ideal for applications requiring velocity and scaling, for instance in manufacturing situations or while handling huge datasets.

Spacy's primary characteristics include the following:

Language support: Spacy handles a number of dialects, including Spanish, German, English, Dutch, French, Italian, and others.

Pre-trained models: Spacy offers models that have been trained for many dialects, which may be uploaded with only a few pieces of script. These representations may be utilised in a variety of tasks involving NLP, including dependency parsing, part-of-speech tagging, named entity recognition, and others.

Tokenization: Spacy uses advanced tokenization techniques to split text into individual words and punctuation marks. It can handle a range of languages and can also split compound words and contractions.

Part-of-speech tagging: Spacy may instantly assign elements of speech, such as a noun, a verb, an adjective, or adverb, to every syllable in a phrase. This may be helpful for a variety of uses including sentiment analysis and text categorization.

Named entity recognition: Spacy is able to recognise and categorise designated entities in written content, such as individuals, groups, places, and occasions, using named entity recognition. This is important for activities like gathering data and object connection.

Dependency parsing: Spacy itself can evaluate the structure of grammar of a phrase and find the links among items via dependency parsing. This is important for activities like analysing sentiment and query response.

Text classification: Spacy includes a range of built-in models for text classification, including sentiment analysis and topic modelling. These models can be trained on custom datasets to create more accurate models for specific use cases.

Customization: Spacy provides a range of tools for customising and training models on specific tasks or domains. This allows developers to create more accurate models for specific use cases and can help improve performance on specific datasets.

Overall, Spacy is a powerful and flexible library for natural language processing in Python. Its rapidity and flexibility render it suitable for usage in commercial situations, and its variety of functions and customizable possibilities make it an appealing option among academics and engineers who are developing an extensive variety of applications that use NLP.

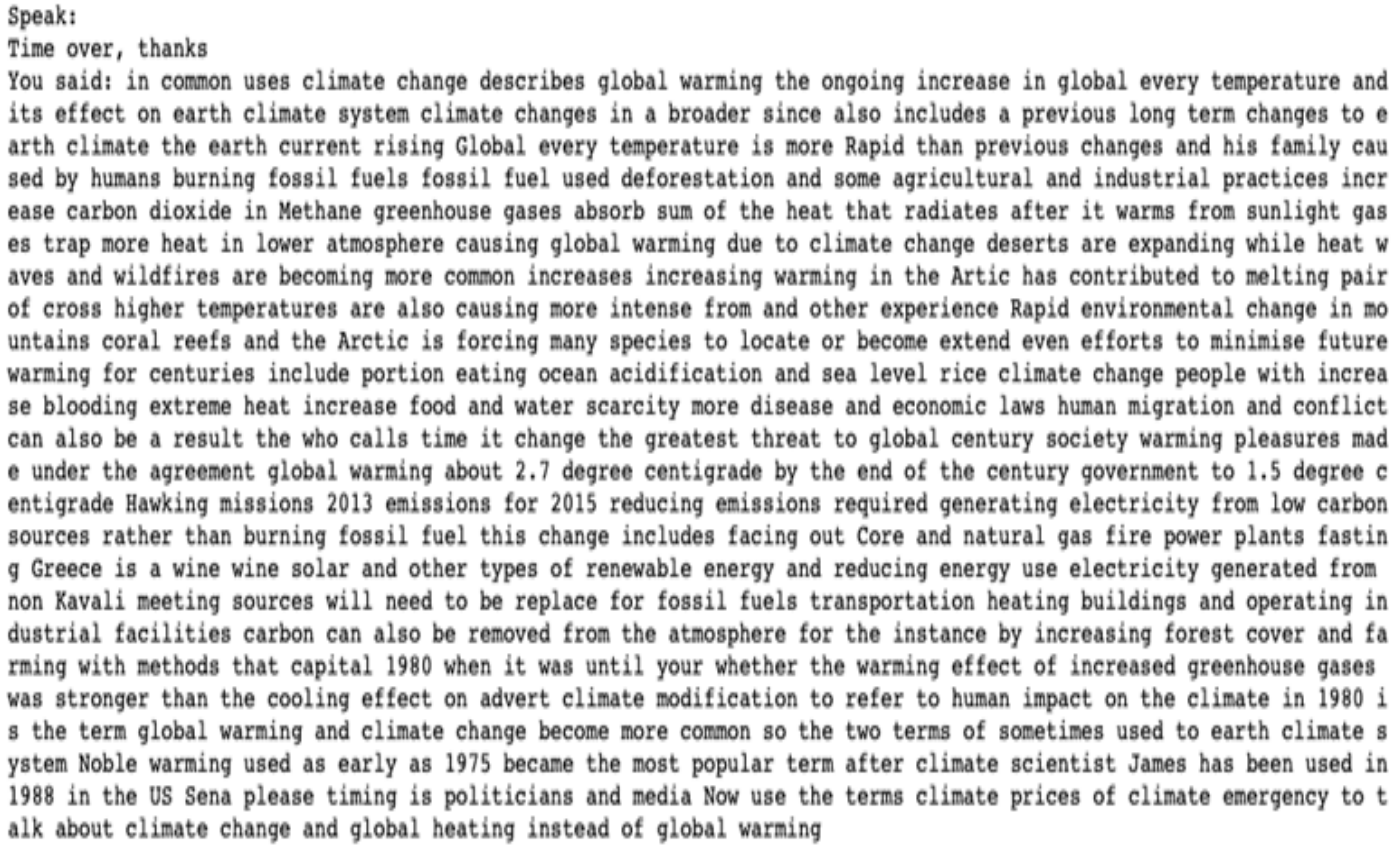


Fig 1 Input Given by the User

ISSN No:-2456-2165

2013 emissions for 2015 reducing emissions required generating electricity from low carbon sources rather than burni ng fossil fuel this change includes facing out Core and natural gas fire power plants fasting Greece is a wine wine s olar and other types of renewable energy and reducing energy use electricity generated from non Kavali meeting source s will need to be replace for fossil fuels transportation heating buildings and operating industrial facilities carbo n can also be removed from the atmosphere for the instance by increasing forest cover and farming with methods that c apital 1980 when it was until your whether the warming effect of increased greenhouse gases was stronger than the coo ling effect on advert climate modification to refer to human impact on the climate in 1980 is the term global warming and climate change become more common so the two terms of sometimes used to earth climate system Noble warming used a s early as 1975 became the most popular term after climate scientist James has been used in 1988 in the US Senaevery temperature is more Rapid than previous changes and his family caused by humans burning fossil fuels fossil fuel used deforestation

Fig 2 Processed Output in the form of Summary

➢ *Block Diagram:*

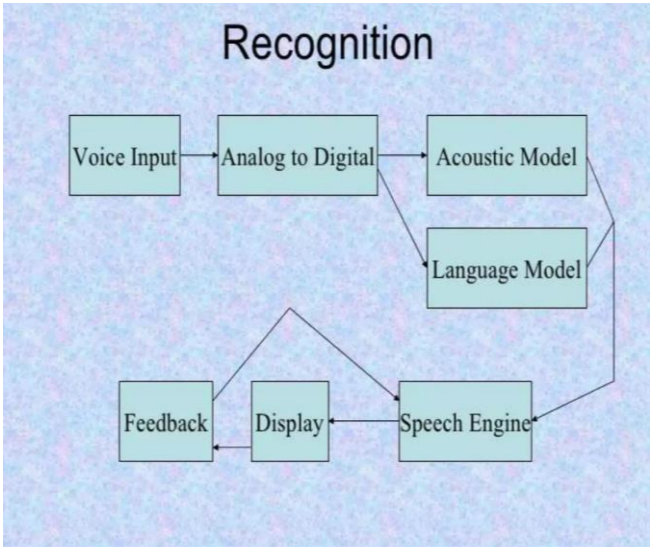


Fig 3 Block Diagram of a Speech Recognition System

Figure 3 depicts the block diagram of a speech recognition system and Figure 4 depicts the activity diagram of a speech recognition system.

➢ *Activity Diagram:*

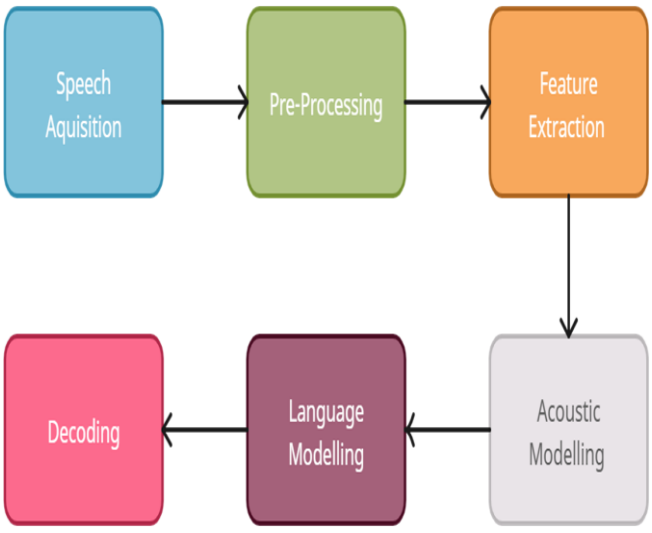


Fig 4 Activity Diagram of a Speech Recognition System

## V. CONCLUSION

In conclusion, meeting insights summarization using speech recognition and summarization techniques presents a promising solution for improving the efficiency and effectiveness of meetings. The use of ASR and summarization technology can provide attendees with a concise and manageable summary of meeting content, allowing them to more effectively retain and recall key information and insights. This paper provides a comprehensive overview of the current state of speech recognition and summarization technology and demonstrates how these technologies can be applied to meeting insights summarization. In our attempt to design the code for a speech summarization system for meetings, we tried using Spacy and Genism libraries to implement the system, and Figure 1 depicts the speech spoken by the user which is processed using the system created by us and Figure 2 displays the processed output in the form of the spoken speech summary. More study is required to investigate the possible advantages and disadvantages of this strategy, in addition to developing more advanced summary algorithms.

Speech recognition is a fast-expanding technology with the possibility to transform how we communicate with machines and other objects. Notwithstanding ongoing obstacles, developments in machine learning and signal processing are enabling the creation of increasingly precise and trustworthy voice recognition networks, which have the ability to alter a broad spectrum of industry sectors and applications.

## REFERENCES

[1]. Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In the European conference on computer vision (pp. 483-499). Springer, Cham.

[2]. Furui, S., Iwano, K., Hori, C., Shinozaki, T., Saito, Y., & Tamura, S. (2001, May). Ubiquitous speech processing. In 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221) (Vol. 1, pp. 13-16). IEEE.

[3]. Furui, S. (2003). Recent advances in spontaneous speech recognition and understanding. In ISCA &

[4]. IEEE workshop on spontaneous speech processing and recognition.
[5]. Hori, C., & Furui, S. (2001). Advances in automatic speech summarization. RDM, 80, 100.
[6]. Furui, S., & Kawahara, T. (2008). Transcription and distillation of spontaneous speech. Springer Handbook of Speech Processing, 627-652.
[7]. Sakshi Bhalla, Roma Verma, Kusum Madaan, 2017, Comparative Analysis of Text Summarisation Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICCCS – 2017 (Volume 5 – Issue 10),
[8]. Radev, D., Hovy, E., & McKeown, K. (2002). Introduction to the special issue on summarization. Computational linguistics, 28(4), 399-408.
[9]. Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023.
[10]. Wan, X., Li, C., Wang, R., Xiao, D., & Shi, C. (2018). Abstractive document summarization via bidirectional decoder. In Advanced Data Mining and Applications: 14th International Conference, ADMA 2018, Nanjing, China, November 16–18, 2018, Proceedings 14 (pp. 364-377). Springer International Publishing.
[11]. https://www.sciencedirect.com/topics/engineering/speech-recognition
[12]. https://pypi.org/project/gensim/
[13]. https://spacy.io/