Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Feature Extraction and Analysis of Speech Signal Using Empirical Mode Decomposition

Mrs.P Madhavi Chowdary[1], J Jayalakshmi[2], L Bharath Babu[3], B Anil Kumar[4], C Dhanasekhar[5], P Imamuddin[6]

(AssistantProfessor, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India)1

(B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India) 2

(B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India) 3
(B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India) 4
(B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India) 5
(B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India) 6
(Email.id: -jayajanyavula128@gmail.com)

--************************----------------------------------

Abstract:

The use of voice, an analogue signal with specific information, to gather and transmit information in social interactions has become more prevalent. Transforming complicated speech settings into useful speech information is the aim of speech signal processing. Over time, the speech signal's characteristics change significantly. An important first step in speech signal processing is to guarantee that the qualities of the voice signal are mostly unchanged in a short amount of time. The accuracy and robustness of feature extraction from speech signals also have an impact on the rate of speech recognition. Speech signal feature extraction is therefore essential in applications involving speech signal processing. Another way to understand the distinct benefits of the Empirical Mode Decomposition (EMD) over Intrinsic Mode Functions (IMFs) is to contrast its time-frequency resolution with that of the fractional Fourier transform on a noisy background.

Keywords **—**Empirical Mode Decomposition (EMD), Speech Signal, Feature Extraction, Speech Recognition, Speech Signal Processing.

--************************----------------------------------

I. INTRODUCTION

In the rapidly evolving field of communication systems, voice signals serve as a crucial medium for conveying and receiving information. Yet, the inherent complexity of speech signals poses a significant challenge in extracting meaningful information. This project is dedicated to exploring the realms of speech signal processing, particularly focusing on the innovative technique of Empirical Mode Decomposition (EMD) for feature extraction. The process of signal processing plays a pivotal role across various sectors, employing techniques to manipulate, analyse, and interpret signals to derive valuable data or enhance signal quality for further use. One of the primary challenges in signal processing is the handling of non-stationary signals—those whose statistical properties change over time. Traditional methods often fall short in capturing the dynamic characteristics of such signals, thus paving the way for advanced methodologies like EMD, which decomposes non-linear and non-stationary signals into intrinsic mode functions (IMFs) for detailed analysis.

The scope of this project encompasses a thorough investigation of EMD's effectiveness in speech signal processing. By decomposing speech signals into their intrinsic mode functions, EMD allows for a deeper understanding of the signal's underlying characteristics. The project will compare EMD with traditional methods such as the Fractional Fourier Transform, assessing its proficiency in handling noisy environments and extracting relevant features crucial for speech recognition. The implementation phase involves integrating EMD into existing speech recognition systems and evaluating its performance through various metrics such as accuracy, robustness, and computational efficiency. This comparative analysis aims to highlight the distinct advantages of using EMD over conventional methods, especially in dynamic speech environments.

Our objectives are to understand EMD's operational intricacies, adapt it to signal processing challenges, and evaluate its practical effectiveness. The project will also delve into signal classification and analysis, employing machine learning algorithms for tasks like pattern recognition and anomaly detection using features extracted via EMD. Furthermore, the exploration extends to assessing EMD's potential applications in broader domains such as telecommunications and biomedical engineering. By evaluating both the technical aspects and practical implications of EMD-based techniques, this project aims to bridge theoretical research with real-world applications, fostering innovation and advancing methodologies in signal processing. Ultimately, the insights gained from this study are expected to contribute to the ongoing evolution of speech recognition technologies and their applications in diverse industries.

II. LITERATURE SURVEY

Authors Li Hong, Xu Xiaoli, Wu Guoxin, et al. [1] present their research in the paper titled "Research on Speech Emotion Feature Extraction Based on MFCC." This study addresses the critical aspect of feature extraction in speaker recognition, highlighting a novel approach that enhances the traditional Mel Frequency Cepstral Coefficients (MFCC). The researchers conducted their experiments using a database of 30 speakers, equally divided between males and females, all recorded in a soundproof room. Their findings demonstrate that the improved MFCC-derived parameters outperform those obtained through traditional MFCC methods utilizing Hidden Markov Models. This advancement underscores the potential for more accurate representation of speaker personality traits in speech signal processing.

Authored by Zhang Hongbing[2] the study titled "Simulation of Speech Signal Deep Spectrum Feature Extraction Method at Mel Frequency" explores the advancements in Mel Frequency Cepstral Coefficients (MFCCs), a cornerstone in speaker and speech recognition technology since the 1980s. This paper delves into the enhancement techniques of MFCCs, crucial for improving the performance and robustness of speech recognition systems. It covers various aspects including the choice of spectral estimation methods, the design of effective filter banks, and the optimal number of features to use. The study provides a comprehensive overview, presenting data on accuracy, environmental types, data nature, and feature count, which are summarized in detailed tables alongside key references. The benefits and limitations of these enhancement techniques are thoroughly discussed, aiming to foster further development towards more robust, accurate, and less complex MFCC features. This research contributes significantly to ongoing efforts in enhancing deep spectrum feature extraction methods at Mel frequency, as evidenced by its detailed analysis and practical evaluations.

Authors Dong Mei, Liao Yunxia, Liu Haishan [3]. present their research in the paper titled "Simulation of Speech Signal Deep Diapason Feature Extraction System at Mel Frequence." This study delves into the rapidly evolving field of speech signal processing, heightened by advances in artificial intelligence. Focusing on the extraction of crucial speech features, the research elaborates on extracting the fundamental period and Linear Predictive Coding (LPC) parameters, as well as detecting resonance peaks from real-time collected speech signals. The paper provides a thorough exploration of typical algorithms, detailing both their programming implementation and algorithmic results. This work is poised to enhance various applications where speech signals are integral, such as speech compression, recognition, and synthesis, showcasing a comprehensive design for a speech signal feature extraction system based on MATLAB.

Authors Yang Lidong, Gu Yu, and Zhang Ming [4] explore advancements in speech recognition technology in their paper titled "Simulation Research on Speech Signal Feature Selection Optimization Extraction." Speech, a complex motor skill refined in adults who can produce approximately 14 distinct sounds per second through the coordinated action of about 100 muscles, presents unique challenges in speaker recognition. This capability enables systems to identify and verify speakers by analysing speech signals. The paper focuses on optimizing the process of feature extraction, which involves converting speech waveforms into a parametric representation to reduce data rates for further processing and analysis. Key techniques such as Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPC), Linear Prediction Cepstral Coefficients (LPCC), Line Spectral Frequencies (LSF), Discrete Wavelet Transform (DWT), and Perceptual Linear Prediction (PLP) are examined for their reliability and adaptability across various applications. The study underscores that no single method is superior; rather, the choice of feature extraction technique depends on the specific application context. Enhancements to these methods aim to reduce susceptibility to noise, increase robustness, and decrease processing time, contributing to the overall efficacy of speaker recognition systems.

Authors Mei Jianmin, Jia Jide, Zeng Ruili [5] in their paper titled "FRFT Cyclic Frequency Filtering and Weak Gear Fault Feature Extraction," address the challenge of accurately and efficiently extracting feature frequencies from early weak fault vibration signals in rolling bearings. They propose a novel method that integrates Intrinsic Time-scale Decomposition (ITD) with Auto Regression (AR) and Minimum Entropy Deconvolution (MED) techniques. The process begins with the decomposition of the original vibration signal using the ITD algorithm, which isolates proper rotations (PRs) containing the fault feature frequencies. Subsequently, the sample entropy value of each PR is calculated to identify the PR with the highest sample entropy. This PR is then subjected to AR-MED filtering to minimize noise, thereby enhancing the precision of the extracted feature frequencies.

III. PROPOSED SYSTEM

Voice, an Analog signal rich with distinct information, has increasingly become a central medium for communication in social interactions. The overarching objective of speech signal processing is to transform complex speech environments into usable speech information. Given the dynamic nature of speech signals, where characteristics can vary significantly over short periods, a crucial step in speech signal processing is ensuring that these characteristics remain relatively stable in brief intervals. The effectiveness of feature extraction from these signals plays a vital role in the success of speech recognition technologies, making it an indispensable process in various speech signal processing applications.

In proposing a method for feature extraction, the process begins with loading a speech signal. This involves several essential steps aimed at readying the audio data for subsequent analysis and processing. It starts with identifying the source of the speech signal, which could either be stored digitally in formats like WAV or streamed live from sources such as microphones or networks. If the data is stored, it undergoes parsing or decoding to extract vital audio parameters including sample rate, bit depth, and the number of channels. For streaming, the signal is captured in real-time and buffered accordingly. This might be followed by sample rate conversion to meet the requirements of specific processing algorithms. The signal is then converted into a one-dimensional array of audio samples, setting the stage for preliminary processing steps like noise reduction and filtering.

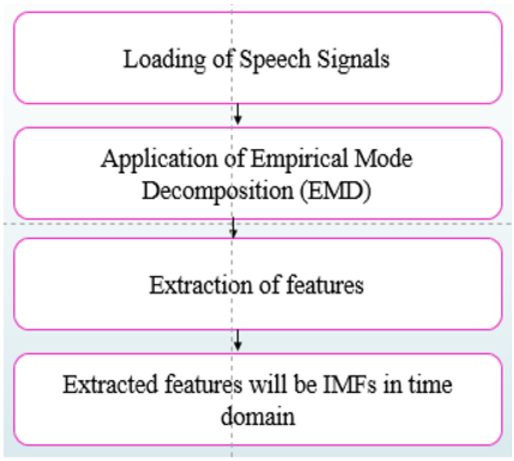

Fig: Flow of Proposed Method

"The process commences by loading the speech data, followed by its decomposition into a series of Intrinsic Mode Functions (IMFs) using EMD. Unlike traditional methods that rely on predefined basis functions, EMD offers a data-driven approach, adapting to the inherent non-stationarities present in speech signals. These IMFs hold the key advantage of capturing localized time-frequency information within the speech. Finally, relevant features are extracted from the IMFs for further analysis. This approach holds significant promise for researchers in speech processing. By exploiting the time-frequency resolution offered by EMD, researchers can potentially gain deeper insights into the characteristics of speech signals, paving the way for advancements in various speech processing applications."

PSEUDO CODE OF THE ALGORITHM

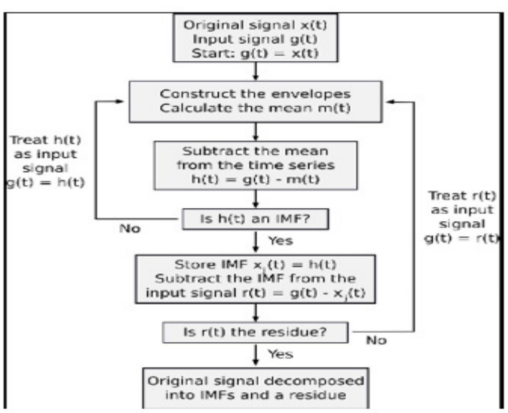

Fig: EMD Algorithm

IV. SOFTWARE DESCRIPTION

MATLAB:

MATLAB, an acronym for Matrix Laboratory, is a sophisticated programming platform and numerical computing environment designed to handle mathematical computations with high efficiency and minimal effort, especially those involving matrices and vectors. Originated in the late 1970s by Cleve Moler, MATLAB was initially created to provide students with easier access to matrix software such as LINPACK and EISPACK without the need to delve into FORTRAN. Over the years, it has evolved significantly from a simple matrix calculator to a powerful tool equipped with a vast array of capabilities including algebraic and differential equation solving, numerical integration, and robust graphical tools for visualizing data in both 2D and 3D. The platform integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in mathematical notation.

Today, MATLAB is not only foundational in academic and educational settings but is also widely employed across various industries for research and development. With its user-friendly interface and expansive toolbox, MATLAB facilitates extensive applications in signal processing, control systems, neural networks, and many other areas of engineering and science. These toolboxes, which are comprehensive collections of MATLAB functions tailored to specific areas, enhance MATLAB's environment to tackle specialized technological tasks efficiently.

V. RESULTS:

Our research successfully utilized Empirical Mode Decomposition (EMD) to analyse and extract crucial features from speech signals. This approach proved highly effective in handling the inherent non-stationarity and variability of speech, enabling the identification of key elements such as pitch, formants, and prosodic features. These findings enhance our understanding of speech dynamics and facilitate more accurate speech processing applications including recognition, identification, and emotion analysis. The adaptability and robustness of EMD underscore its potential as a valuable tool in advanced speech signal analysis, setting a strong foundation for future innovations in the field.

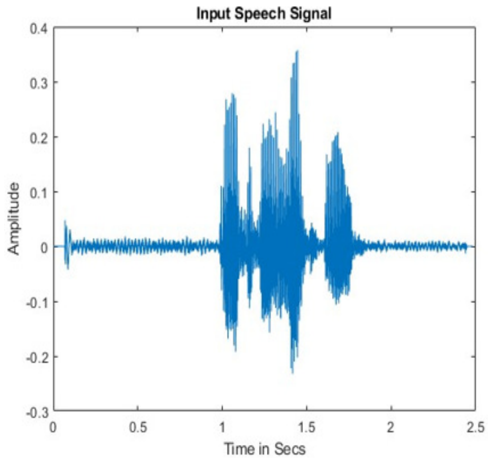

Fig: Input Speech Signal

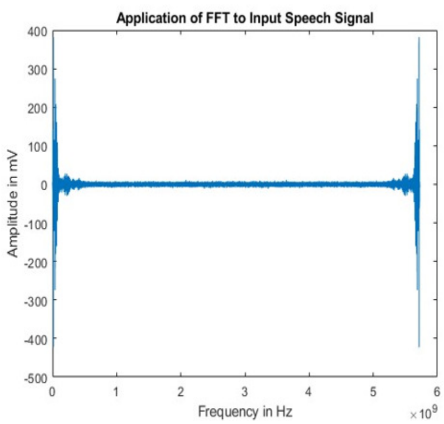

Fig: Input Speech Signal

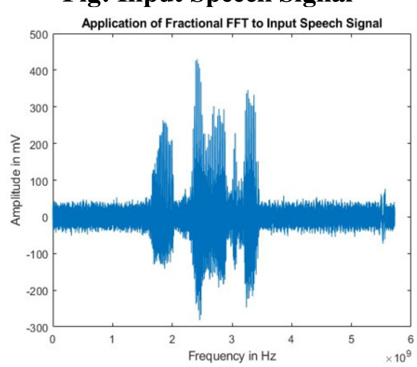

Fig: Application of Fractional FFT to input speech signal

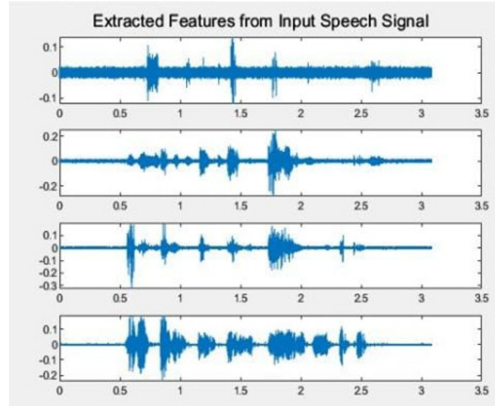

Fig: Extracted features from input speech signal

VI. CONCLUSION

The implementation of Empirical Mode Decomposition (EMD) for feature extraction in speech signal analysis proved effective, demonstrating superior performance over traditional frequency domain features. EMD's ability to decompose signals into Intrinsic Mode Functions (IMFs) allows for detailed analysis of complex, non-stationary speech signals, providing key insights into speech dynamics. Integrating EMD with machine learning enhances tasks like speech recognition and emotion detection. However, optimizing EMD parameters and managing computational demands remain challenges.

Future research will focus on refining EMD parameters and exploring noise reduction techniques to improve robustness in noisy environments. There is potential for integrating EMD with deep learning for advanced speech analysis systems and expanding its application to real-time processing and multimodal data. Further development of EMD can impact various fields, including healthcare and security, promoting innovative speech analysis applications and technologies.

VII. REFERENCES

1. Li Hong, Xu Xiaoli, Wu Guoxin, et al. Research on speech emotion feature extraction based on

MFCC. Journal of Electronic Measurement and Instrument, vol. 31, no. 3, pp. 448-453, 2017.

2. Zhang Hongbing. Simulation of speech signal deep spectrum feature extraction method at Mel frequency. Computer simulation, vol. 37, no. 5, pp. 202-205+272, 2020.

3. Zhang Hongbing. Simulation of speech signal deep spectrum feature extraction method at Mel frequency. Computer simulation, vol. 37, no. 5, pp. 202-205+272, 2020.
4. Yang Lidong, Gu Yu, Zhang Ming. Simulation research on speech signal feature selection optimization extraction. Computer simulation, vol. 33, no. 2, pp. 409-412,417, 2016.

5. Mei Jianmin, Jia Jide, Zeng Ruili, et al. FRFT cyclic frequency filtering and weak gear fault feature extraction. Vibration and shock, vol. 35, no. 23, pp. 130-135, 2016.

6. Yang Lin, Wang Guohong, Yang Zhong. Research on noise feature extraction and classification of radar suppression jamming signalbased on FrFT domain. Ship Electronic Engineering, vol. 37, no. 9, pp. 53-58, 2017.

7. Hao Tianyi, Xue Yonghua, Zheng Zhigang. Feature extraction of noisy signals based on short-time fractional Fourier transform. Journal of Naval Aeronautical Engineering Institute, vol. 34, no. 3, pp. 273- 276, 2019.
8. Yun Tao, Yu Xiang, Wang Jun, et al. ISAR horizontal calibration based on fractional Fourier transform. Data acquisition and processing, vol. 33, no. 1, pp. 106-112, 2018.

9. Anonymous. Application of Fractional Fourier Transform in Chirp Signal Filtering. Ship Electronic Engineering, vol. 12, no. 282, pp. 45-47+51, 2017.

10. Zhang Zhiqiang, Mei Jianmin, Chang Chun, et al. Gear early fault feature extraction based on FRFT target separation and MCKD feature enhancement. Journal of the Academy of Military Transportation, vol. 21, no. 6, pp. 44-48, 2019.