International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

A COMPARISION OF LPC, RASTA, AND MFCC ALGORITHMS IN AN AUTOMATED SPEECH RECOGNITION SYSTEM

[1]K. Siva Pavani, [2]Dr Ch. Venugopal Reddy, [3]A. Aswini, [4]Ch. Kalavarshini, [5]B. Likhitha,[6]K. Mahitha

[2]Professor&HOD, ECE Dept, RISE Krishna Sai Prakasam Group of Institution, Ongole-523001, AP

[1,3,4,5,6]B.Tech final year students, ECE Dept, RISE Krishna Sai Prakasam Group of Institution, Ongole-523001, AP

([1]kokkiligaddasivapavani@gmail.com; [2]phdvenu@gmail.com; [3]aswini9amara@gmail.com; [4]csrcrs44@gmail.com; [5]bodduloorilikhitha28@gmail.com; [6]mahithakancharla09@gmail.com)

ABSTRACT

Speech is a long-standing topic of study that continues to be researched today. The study and recognition of speech signals by machines or computers in a variety of settings is the focus of automatic speech recognition systems. Several feature extraction techniques are used to improve the system's accuracy and functionality. An overview of speech recognition systems and their different stages, including analysis, feature extraction, modelling, testing, and matching, is given in this research study. Furthermore, it comprises a thorough and comparative analysis of the feature extraction methods utilised in Automatic Speech Recognition systems, such as Mel-Frequency Cepstral Coefficient (MFCC), Relative Spectral Filtering (RASTA), and Linear Predictive Coding (LPC). This research paper's primary goal is to provide a concise overview of the speech recognition system and the three feature extraction techniques that are essential to automatic speech recognition. This work presents a novel enhanced approach to speech/speaker recognition that combines the discrete wavelet transform (DWT) with the Relative Spectra Perceptual Linear Prediction (RASTA- PLP) for feature extraction.

INTRODUCTION

The primary goal of a speech/speaker recognition system is to achieve the highest recognition rate accuracy while minimizing the amount of time required for system testing and training. We overcame the aforementioned issues by selecting a feature extraction technique that works well and by implementing neural network classifiers in parallel. This served as the central concept for our work. In order to extract features, the DWT and RASTA-PLP are combined. DWT's multi-resolution, multi-scale analytic features have demonstrated their suitability for processing non-stationary signals, such as speech. On the other hand, RASTA-PLP's resilience to noise is its primary benefit. The objective of this method is to enhance the better performance of the new method in for feature extraction giving a higher recognition rate than using the combination

between DWT and MFCCs based method.

LITERATURE SURVEY

It is well known that speech improvement techniques based on spectral subtraction can effectively reduce additive stationary, wideband noise. The output voice quality is shown to be substantially degraded by tonal stimuli, such as car horn sounds. In order to improve tonal noise suppression, a technique that integrates RASTA processing into the spectral subtraction framework is presented in this study. It is demonstrated that when both broadband and tone disturbances are present at the same time, the suggested approach performs noticeably better than both the spectrum subtraction and RASTA speech enhancement techniques.

Speech parameterization aims to extract from the audio signal the pertinent information about what is being spoken. Relative spectral analysis and Mel-Frequency Cepstral Coefficients (MFCC) are used in voice recognition systems. RASTA-MFCC, or Mel-Frequency Cepstral Coefficients, are the two primary methods employed. This project will demonstrate how it offers various changes to the original MFCC approach. We examined and compared the Modified Function Cepstral Coefficients (MODFCC), which are suggested modifications to MFCC, to the original MFCC and RASTA-MFCC characteristics. Jitter and shimmer are examples of prosodic qualities that are added to baseline spectral features. The aforementioned methods were evaluated using impulsive signals in AURORA databases under a range of noisy situations.

For continuous speech in Myanmar, this project offers automatic speech recognition. In actuality, it is not assumed that a machine or computer will comprehend what is said. However, it is anticipated that speech control or the conversion of the acoustic signal into symbols will be used. Additionally, this method will solve the problem of automatically detecting word/sentence boundaries in loud and quiet settings. In feature extraction, a combination of LPC, MFCC, and GTCC algorithms are employed. The MFCC characteristics provide good speech signal discrimination. LPC is a computational model of speech that is effective and gives a precise estimation of the speech parameters.HMM is utilized for recognition, and DTW is used for feature clustering. The benefits of these two potent pattern recognition techniques are combined by extending the HMM approach by merging it with the DTW algorithm.

voice synthesis, voice noise reduction, and speaker recognition are some of the main research topics in this crucial field. One of the new challenges for technology is speaker recognition. Numerous feature extraction algorithms have been proposed and created. The Mel Frequency Cepstral Coefficients (MFCC) feature extraction method for speaker detection is presented in this research. Additionally, studies carried out at every stage of the MFCC process are evaluated in this study. Lastly, using a Matlab environment, the study contrasts the rectangular window and hamming window techniques based on the number of filters for precise and effective results. The

outcome shows that, in comparison to other windowing strategies and filter counts, employing a 32 filter with a hamming window offers greater accuracy and efficiency.

We demonstrate a throat microphone-based speech recognition system in this research. By using this type of microphone, the effects of background noise are reduced. The identification rate of throat microphone systems has been lower than that of normal microphone systems due to the lack of high frequencies and the partial loss of formant frequencies. We present two approaches to build a throat microphone-only high performance automated speech recognition (ASR) system. Using a thorough analysis of the throat signal and Korean phonological feature theory, we first demonstrate that an ASR system can be created with just a throat microphone. We next provide requirements for the feature extraction algorithm. We suggest a choice of cochlear filters and an increase of the formant frequencies for ZCPA optimization. Comparing this system to the performance of a standard ZCPA algorithm on throat microphone signals, experimental results indicate that it improves performance by around 4% and reduces time complexity by 25%.

The field of automatic speech recognition (ASR) is fascinating, and a lot of study has been done in this area by numerous researchers. Additionally, development in digital signal processing has made it possible for computer hardware to speak human language. In addition to all of these developments and studies in digital signal processing, computer systems still cannot equal human speech in terms of matching accuracy and response time. Currently, the goal of the speech recognition process is to develop an ASR system that is independent of the speaker. the causes of its many applications and the limitations of the automatic speech recognition methods now in use.

This study demonstrates the use of speech recognition for arm robot control. Linear Predictive Coding (LPC) and the Adaptive Neuro-Fuzzy Inference System (ANFIS) are used in this technique to identify speech recognition. The ANFIS approach is utilized to train speech recognition, whereas the LPC method is used to feature extract the speech signal. Six features are utilized in the data learning process that ANFIS processes. Both trained and untrained data are used in the speech identification examination system. The study's findings indicate that the successful grade for trained speech data is 88.75%, while the untrained data is 78.78%. A speech recognition system was implemented on an Arduino microcontroller-based controlled arm robot.

Existing System

Suggests employing wavelet-based MFCC to improve speech identification. This involves applying the wavelet transform before conventional MFCC in order to record more intricate speech characteristics at various resolutions. Better time-frequency resolution is made possible

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

by the wavelet transform, which enables the system to record both rapid and slow speech changes, enhancing the understanding of subtleties like intonation and pitch.

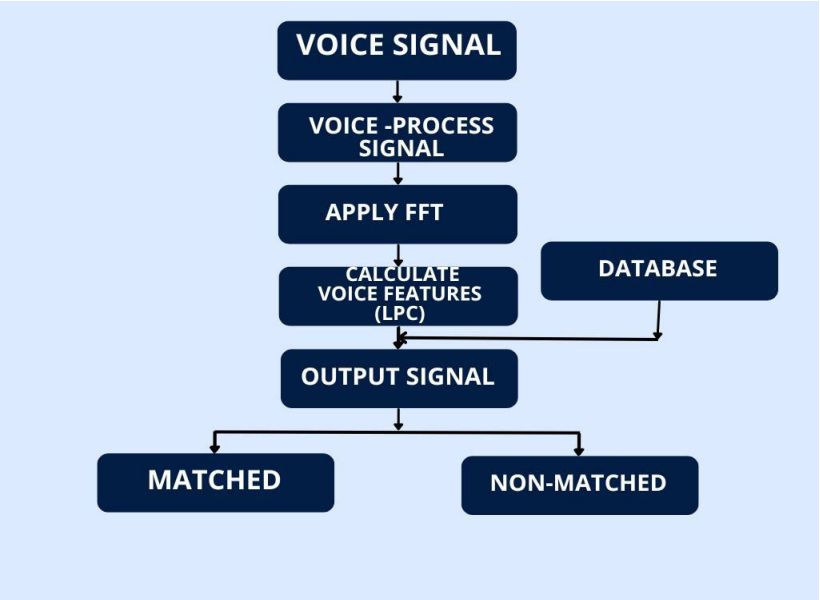

Fig 1: Existing System block diagram

1.**Voice Signal Acqaisation:** Capturing of the voice signal.capturing of the voices are

Direct way

Pre-Defined way

In this project we are using predefined method. These voice signals are by using

i=uigetfile('*.wave');

j=waveread(i)

2.Pre-emphasis of the voice Signal:

Comprehension voice signal

Voice signal compression involves reducing the size of audio data for efficient transmission or storage, typically by removing redundant or unnecessary information. In the context of students, it ensures faster download times, better use of bandwidth, and clearer communication in educational applications or online classes.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

De-Noise of the voice signal

De-noising the voice signal involves removing background noise to enhance the clarity and quality of the audio. This process is achieved through techniques such as filtering or advanced machine learning algorithms to ensure a cleaner, more intelligible sound.

3.FFT : The fast fourier transform is an algorithm used to efficiently compute the discrete fourier transform.

The DFT has a direct computation complexity of O(N^2), where N is a no.of points in the sequence. The FFT reduces this complexity to O(N logN).

Finally, we can calculate time or frequency features.

4.Voice Features : Voice features are mean, variance, standard deviation.

5. Matched/Non-Matched: The comparision determines whether the input voice matches any of the stored templates, resulting in either a "Matched" or "Non-Matched" output.

Proposed System

We suggested speaker/speech recognition. Perceptual linear prediction, or RASTA-PLP, is the technique we employ in this suggested approach. Hermansky invented this method. By employing a high pass filter to eliminate the slowly fluctuating components in each element of the filter bank output, RASTA can overcome the issues brought on by the communication routes.

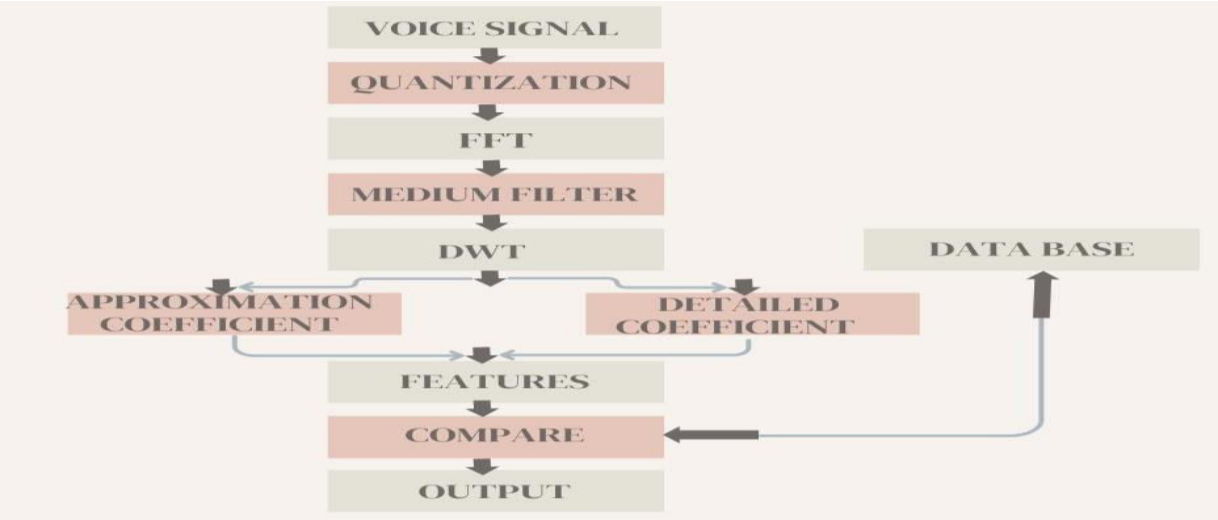

Fig 2: Proposed System block Diagram

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

1. Voice Signal: Start with recording or collecting a waveread (to read audio or voice) signal.

2. Quantization: Convert the recorded sound into digital form so it can be processed by a computer.

3. FFT : Break the sound into its frequency components to better analyze its properties.

4. Medium Filter: Clean the signal by removing unwanted noise or distortions.

5. DWT (Discrete Wavelet Transform): In DWT, the original signal passes through two filters

Low-pass filter producing the approximation coefficients h[n] which is the most important part.

High-pass filter which produce the detail coefficients g[n]

Here two functions, called scaling function and wavelet function, are used in finding the DWT are

$$\phi(t) = \sum_{n=0}^{N-1} h(t)\sqrt{2}\phi(2t - n)$$

$$\psi(t) = \sum_{n=0}^{N-1} g(n)\sqrt{2}\phi(2t - n)$$

A two stage filter process is shown in figure where A denotes the approximation coefficients and D denotes the detail coefficients.

The process of down-sampling is that what produces the DWT coefficients which are the approximation coefficients obtained from the low pass filter and the detail coefficients obtained from the high pass filter.

A multi level wavelet decomposition tree is obtained by further decomposition process for the approximation coefficients, so that one signal is broken down into many lower resolution components. where the signals after decomposition will be as

S= cA3 + cD3 + cD2 + cD1

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

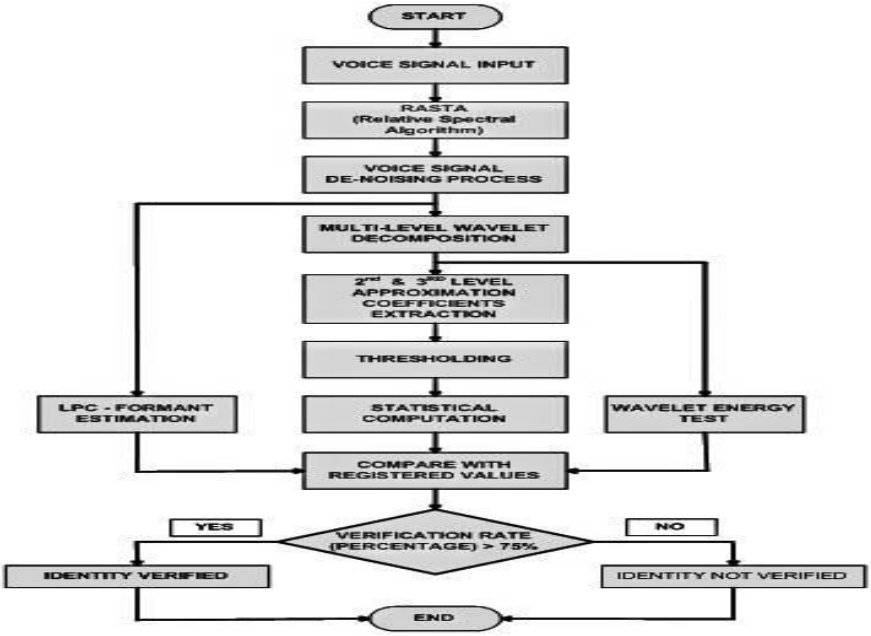

Fig 3: System Implementation

Results

The GUI for the "Speaker Registration Section" guides users through selecting and processing audio files. It includes steps for selecting a file, denoising the signal, and analyzing statistical measures. The interface displays coefficients, approximations, and format values like formants and wavelet energy for speaker registration

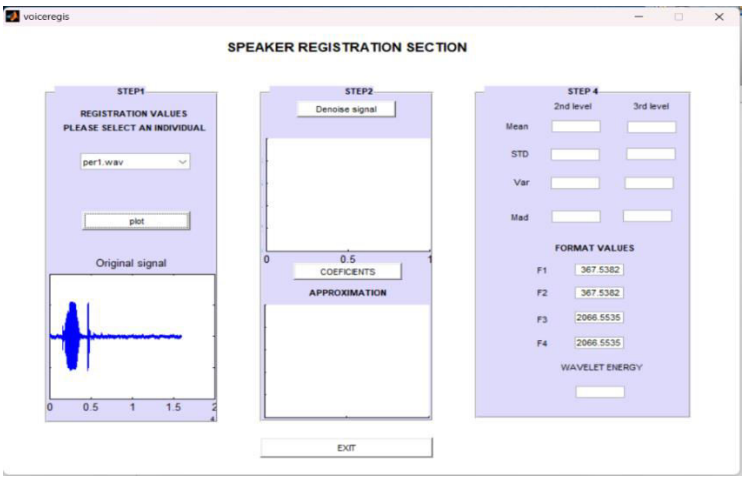

Fig 4: Speaker registration section

The "Speaker Registration Section" GUI allows users to select an audio file, plot the original

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

signal, and proceed with signal denoising. It provides statistical analysis at different levels and displays format values like formants and wavelet energy, helping in speaker feature extraction.

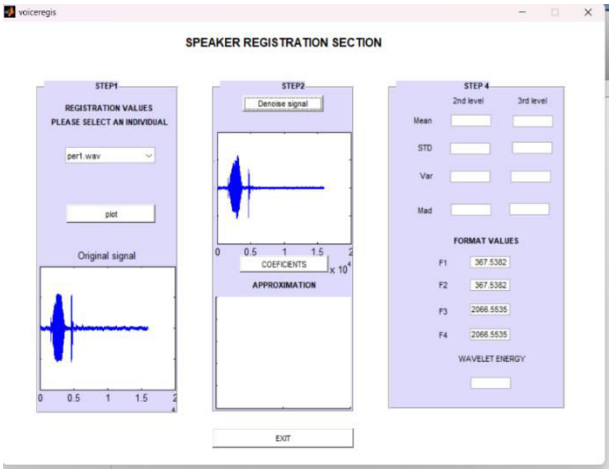

Fig 5: Improvement in output

The "Speaker Registration Section" GUI allows users to select an audio file, plot the original signal, and proceed with signal denoising. It provides statistical analysis at different levels and displays format values like formants and wavelet energy, helping in speaker feature extraction.It shows a section with steps to register a speaker's voice, including: Selecting an individual, Denoising the signal, Extracting coefficients and features, and Formatting the values. The VERIFICATION panel shows the verification process of the system. The overall percentage values of the statistical computation, formant values and the wavelet energy are displayed.

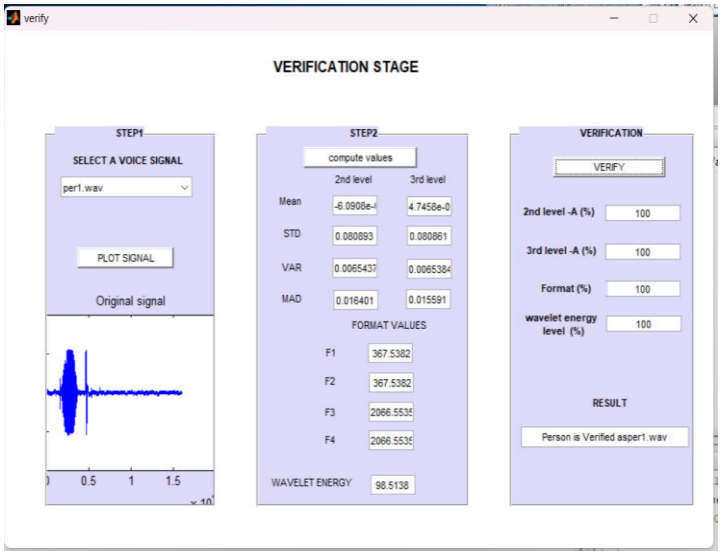

Fig 6: Verification step

CONCLUSION

In order to do speech recognition, LPC (Linear Predictive Coding) is used to estimate the formant and determine the pitch of the voice stream. The created speech recognition system is a word-dependent voice verification system that uses wavelet energy, formant estimation, and statistical calculation to confirm a person's identity based on their own voice signal. Verification tests have been conducted utilising fifty preloaded speech signals from five different people, and an accuracy rate of roughly 80% has been attained.

REFERENCES

[1] Soontorn Oraintara, Ying-Jui Chen Et.al. IEEE Transactions on Signal Processing, IFFT, Vol. 50,No. 3, March 2002

[2] Kelly Wong, Journal of Undergraduate Research, The Role of the Fourier Transform in Time-Scale Modification,University of Florida, Vol 2, Issue 11 - August 2001

[3] Bao Liu, Sherman Riemenschneider, An Adaptive Time- Frequency Representation and Its Fast Implementation, Department of Mathematics, West Virginia University

[4] Viswanath Ganapathy, Ranjeet K. Patro, Chandrasekhara Thejaswi, Manik Raina, Subhas K. Ghosh, Signal Separation using Time Frequency Representation, Honeywell Technology Solutions Laboratory

[5] Amara Graps, An Introduction to Wavelets, Istituto di Fisica dello Spazio Interplanetario, CNR-ARTOV

[6] Brani Vidakovic and Peter Mueller, Wavelets For Kids – A Tutorial Introduction, Duke University

[7] O. Farooq and S. Datta, A Novel Wavelet Based Pre Processing For Robust Features In ASR

[8] Giuliano Antoniol, Vincenzo Fabio Rollo, Gabriele Venturi, IEEE Transactions on Software Engineering, LPC & Cepstrum coefficients for Mining Time Variant Information from Software Repositories

[9] Michael Unser, Thierry Blu, IEEE Transactions on Signal Procesng, Wavelet Theory Demystified,Vol. 51,No. 2,Feb'03

[10] Viswanath Ganapathy, Ranjeet K. Patro, Chandrasekhara Thejaswi, Manik Raina, Subhas K. Ghosh, Signal Separation using Time Frequency Representation, Honeywell Technology Solutions Laboratory

[11] Amara Graps, An Introduction to Wavelets, Istituto di Fisica dello Spazio Interplanetario, CNR-ARTOV

[12] Brani Vidakovic and Peter Mueller, Wavelets For Kids – A Tutorial Introduction, Duke University

[13] O. Farooq and S. Datta, A Novel Wavelet Based Pre Processing For Robust Features In ASR

[14] Giuliano Antoniol, Vincenzo Fabio Rollo, Gabriele Venturi, IEEE Transactions on Software Engineering, LPC & Cepstrum coefficients for Mining Time Variant Information from Software Repositories

[15] Michael Unser, Thierry Blu, IEEE Transactions on Signal Procesng, Wavelet Theory Demystified,Vol. 51,No. 2,Feb'03