Advanced Speaker Verification System Using Wavelets

G.Madhavilatha , A.Mallaiah, T.Venkata Lakshmi

[1] P.G Student in ECE Dept., GEC, Gudlavalleru,
[2] Associate Professor in ECE Dept.,GEC, Gudlavalleru
[3] Associate Professor in ECE Dept.,GEC,Gudlavalleru

Abstract—

This paper presents an advanced speaker verification system using wavelet transforms. The conventional signal processing techniques assume the signal to be stationary and are ineffective in recognizing non stationary signals such as voice signals. Voice signals which are more dynamic could be analysed with far better accuracy using wavelet transform. The developed voice verification system is *word independent* voice verification system combining the RASTA and LPC. The voice signal is filtered using the special purpose voice signal filter using the Relative Spectral Algorithm (RASTA).The signals are de-noised and decomposed to derive the wavelet coefficients and there by a statistical computation is carried out. Further the formant of the resonance of the voices signals is detected using the Linear Predictive Coding (LPC).

Keywords- LPC, RASTA, Voice verification model

I. INTRODUCTION

Voice signal is becoming one of the major parts in human's daily lives. The basic function of the voice signal is that it is used as one of the major tools for communication. However, due to technological advancement, the voice signal is further processed using software applications and the information is utilized in various application systems.

The fundamental idea of this project is to use wavelets as a mean of extracting features from a voice signal. The wavelet technique is considered a relatively new technique in the field of signal processing compared to other methods or techniques currently employed in this field. Current methods used in the field of signal processing include Fourier Transform and Short Term Fourier Transform (STFT) [1] [2]. However due to severe limitations imposed by both the Fourier Transform and Short Term Fourier Transform in analyzing signals deems them ineffective in analyzing complex and dynamic signals such as the voice signal [3][4]. In order to substitute the short-comings imposed by both the common signal processing methods, the wavelet signal processing technique is used.

The wavelet technique is used to extract the features in the voice signal by processing data at different scales. The wavelet technique manipulates the scales to give a higher correlation in detecting the various frequency components in the signal. These features are then further processed in order to construct the voice recognition system. Extracting the features of the voice signal does not limit the capabilities of this technique to a particular application alone, but it opens the door to a wide range of possibilities as different applications can benefit from the voice extracted features.

Applications such as speech recognition system, speech to text translators, and voice based security system are some of the future systems that can be developed.

II. WAVELET TRANSFORM

Wavelets are mathematical functions that satisfy certain requirements **[5].** From a mathematical point of view, the wavelet is described as a function that should integrate to zero and it has a waveform that has a limited duration. The wavelet is also finite of length which means that it is compactly supported. Wavelets analyze a signal using different scales. This approach towards signal processing is called multi-resolution analysis. The scale is similar to the window function in STFT.

Multi-Resolution Analysis (MRA) analyze the frequency components of the signal with different resolutions. This approach especially makes sense for non-periodic signal such as the voice signal which has low-frequency components dominating for long durations and short durations of high-frequency components [6]. A large scale can be interpreted as a “large” window. Using a large scale to analyze the signal, the gross features of a signal can be obtained. Vice versa, a small scale is interpreted as a “narrow” window and the small scale is used to detect the fine details of the signal. This property of wavelet analysis makes it very powerful and useful in detecting or revealing hidden aspects of data and since wavelet transform provides a different perspective in analyzing a signal, compression or de-noising a signal can be carried out without much signal degradation. Local features of a signal can be detected with far better accuracy with wavelet transform **[**7**]**.

A. Discrete Wavelet Trasfrom

DWT compensates for the huge amount of data generated by the CWT. The basic operation principles of DWT are similar to the CWT however the scales used by the wavelet and their positions are based upon powers of two. This is called the dyadic scales and positions as the term dyadic stands for the factor of two [9]. For a voice signal, if the high frequency content is removed, the voice will sound different but the message can still be heard or conveyed. This is not true if the low frequency content of the signal is removed as what is being spoken cannot be heard except only for some random noise.

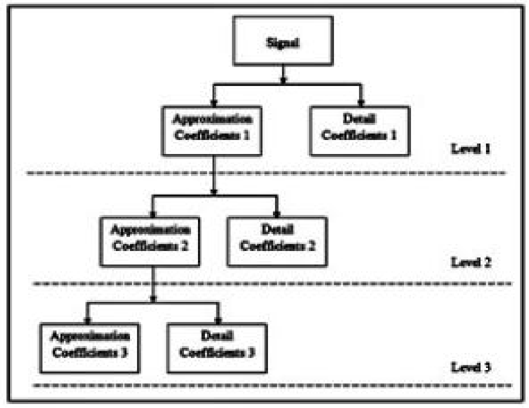

Fig 1. DWT Operation

The DWT is represented using the mathematical equation

$$C(a,b) = C(j,k) = \sum_{n=1} f(n)\Psi j,k(n)_{.1}$$

Where the discrete wavelet $\Psi_{j,k}$ is described as:

$$\Psi j,k(n) = Z^{-1/n}\Psi(Z^{-1}n - k).$$

The basic operation of the DWT is that the signal is passed through a series of high pass and low pass filter to obtain the high frequency and low frequency contents of the signal. The low frequency contents of the signal are called the approximations. This means the approximations are obtained by using the high scale wavelets which corresponds to the low frequency. The high frequency components of the signal called the details are obtained by using the low scale wavelets which corresponds to the high frequency.

From Figure 1, demonstrates the single level filtering using DWT. First the signal is fed into the wavelet filters. These wavelet filters comprises of both the high-pass and low-pass filter. Then, these filters will separate the high frequency content and low frequency content of the signal. However, with DWT the numbers of samples are reduced according to dyadic scale. This process is called the sub-sampling. Sub-sampling means reducing the samples by a given factor. Due to the disadvantages imposed by CWT which requires high processing power the DWT is chosen due its simplicity and ease of operation in handling complex signals such as the voice signal.

B. Wavelet Energy

Whenever a signal is being decomposed using the wavelet decomposition method, there is a certain amount or percentage of energy being retained by both the approximation and the detail. This energy can be obtained from the wavelet bookkeeping vector and the wavelet decomposition vector. The energy calculated is a ratio as it compares the original signal and the decomposed signal.

III. VOICE SIGNAL ANALYSIS

A. RASTA(Relative Spectral Algorithm)

RASTA or Relative Spectral Algorithm as it is known is a technique that is developed as the initial stage for voice recognition. This method works by applying a band-pass filter to the energy in each frequency sub-band in order to smooth over short-term noise variations and to remove any constant offset. In voice signals, stationary noises are often detected. Stationary noises are noises that are present for the full period of a certain signal and does not have diminishing feature. Their property does not change over time. The assumptions that need to be made is that the noise varies slowly with respect to speech. This makes the RASTA a perfect tool to be included in the initial stages of voice signal filtering to remove stationary noises. The stationary noises that are identified are noises in the frequency range of 1Hz - 100Hz.

B. Formant Estimation

Formant is one of the major components of speech. The frequencies at which the resonant peaks occur are called the formant frequencies or simply formants. The formant of the signal can be obtained by analyzing the vocal tract frequency response.

Figure 2 shows the vocal tract frequency response. The x-axis represents the frequency scale and the y-axis represents the magnitude of the signal. As it can be seen, the formants of the signals are classified as F1, F2, F3 and F4. Typically a voice signal will contain three to five formants. But in most voice signals, up to four formants can be detected. In order to obtain the formant of the voice signals, the LPC (Linear Predictive Coding) method is used. The LPC (Linear Predictive Coding) method is derived from the word linear prediction. Linear prediction as the term implies is a type of mathematical operation. This mathematical function which is used in

discrete time signal estimates the future values based upon a linear function of previous samples [8].

$$\hat{X}(n) = -\sum_{l=1}^{p} a_l x(n-l) \qquad (3)$$

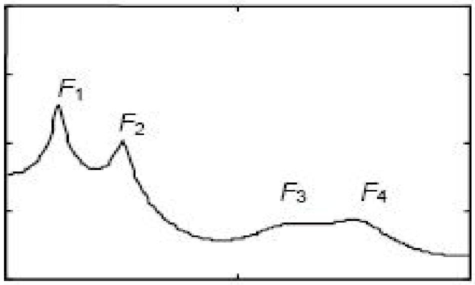

Fig 2.Vocal tract Frequency Response

Where $\hat{X}(n)$ the predicted or estimated is value and $X(n-l)$ is the pervious value.

By expanding this equation,

$$f(n) = -a(1)x(n-1) - a(2)x(n-2).....$$

The LPC will analyze the signal by estimating or predicting the formants. Then, the formants effects are removed from the speech signal. The intensity and frequency of the remaining buzz is estimated. So by removing the formants from the voices signal will enable us to eliminate the resonance effect.

This process is called inverse filtering. The remaining signal after the formant has been removed is called the residue. In order to estimate the formants, coefficients of the LPC are needed. The coefficients are estimated by taking the mean square error between the predicted signal and the original signal. By minimizing the error, the coefficients are detected with a higher accuracy and the formants of the voice signal are obtained.

IV. SYSTEM IMPLEMENTATION

A. *Variation System Implementation*

The proposed methodology for the recognition phase is the statistical calculation. Four different types of statistical calculations are carried out on the coefficients. The statistical calculations that are carried out are mean, standard deviation, variance and mean of absolute deviation. The wavelet that is used for the system is the symlet 7 wavelet as that this wavelet has a very close correlation with the voice signal. This is determined through numerous trial and errors. The coefficients that are extracted from the wavelet decomposition process is the second level coefficients as the level two coefficients contain most of the correlated data of the voice signal. The data at higher levels contains very little amount of data deeming it unusable for the recognition phase. Hence for initial system implementation, the level two coefficients are used.

The coefficients are further threshold to remove the low correlation values, and using this coefficients statistical computation is carried out. The statistical computation of the coefficients is used in comparison of voice signal together with the formant estimation and the wavelet energy. All the extracted information acts like a 'fingerprint' for the voice signals. The percentage of verification is calculated by comparing the current values signal values against the registered voice signal values.

The percentage of verification is given by:

Verification % = (Test value / Registered value) x100 (5)

Between the tested and registered value, whichever value is higher is taken as the denominator and the lower value is taken as the numerator.

Figure 3 shows the complete flowchart which includes all the important system components that are used in the voice verification program.

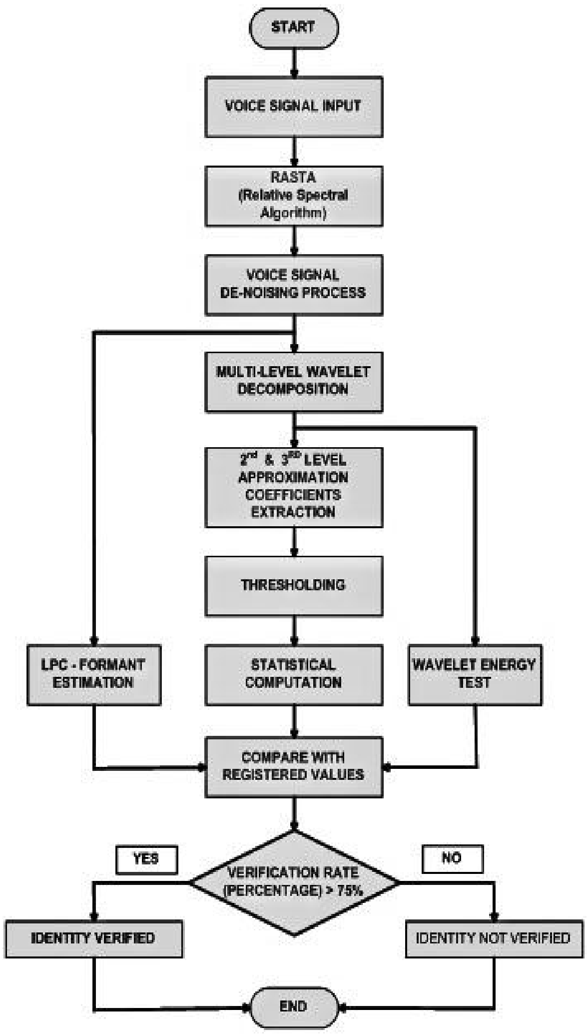

Fig 3.complete System Flowchart.

B. *GUI(Graphical User Interface)*

Figure 4 shows the GUI (Graphical User Interface) implementation for the Voice Registration Section. This GUI enables the user to register an individual's voice signal using the pre-loaded voice signals that are saved in the program.

The STEP 1 panel shows the pre-loaded voice signals contained in the program. The voice signals are obtained from a clean and noise-free environment. The user can select the available voice signal from the popup menu. The Show Voice Plot button enables the user to view the voice plot in the graph shown in the panel.

The STEP 2 panel contains the function that enables the user to de-noise the signal. By pressing the De-Noise button, the user will be able to de-noise the signal and view the de-noised signal in the graph shown in the plot. The Extract

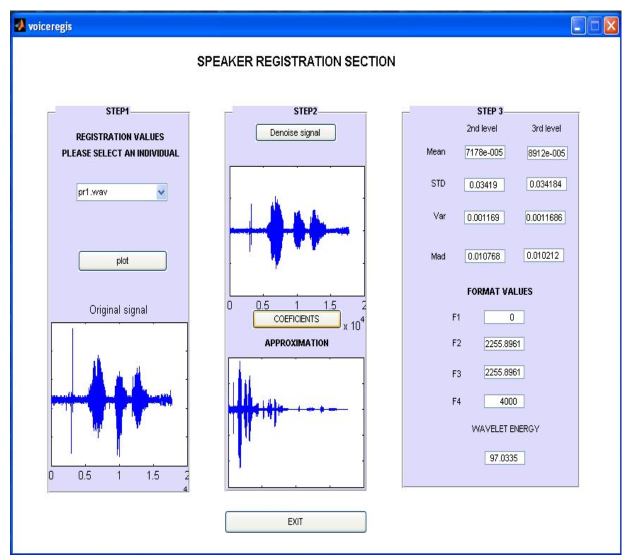

Fig 4. Voice Registration GUI

Coefficients button enables the user to view the DWT (Discrete Wavelet Transform) Coefficients Detail and Approximation.

The STEP 3 panel shows the recognition methodology of the program. The Compute Values button performs the statistical computation on the 2nd level approximation and the 3rd level approximation and displays these values. At the same time the formant values and the wavelet energy of the voice signal is calculated and shown.

Figure 5 shows the GUI (Graphical User Interface) implementation for the Voice Verification Section. This GUI enables the user to verify an individual's voice signal using the pre-loaded voice signals that are saved in the program.

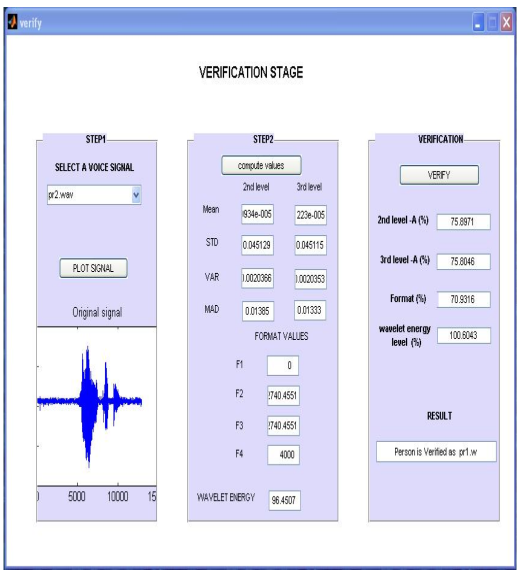

Fig 5. Voice Verification GUI

The STEP 1 panel shows the pre-loaded voice signals contained in the program. The user can select the available voice signal from the pop-up menu. The Plot Voice Signal button enables the user to view the voice plot in the graph shown in the panel. The STEP 2 panel shows the recognition methodology of the program. The Compute Values button performs the statistical computation of the 2nd level approximation and the 3rd level approximation and displays these values. At the same time the formant values and the wavelet energy of the voice signal is calculated and shown. 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.

V. Conclusion

The Speaker Recognition Using Wavelet Feature Extraction employs wavelets in voice recognition for studying the dynamic properties and characteristics of the voice signal. This is carried out by estimating the formant and detecting the pitch of the voice signal by using LPC (Linear Predictive Coding). The voice recognition system that is developed is word independent voice verification system used to verify the identity of an individual based on their own voice signal using the statistical computation, formant estimation and wavelet energy. A GUI is built to enable the user to have an easier approach in observing the step-by-step process that takes place in Wavelet Transform.

REFERENCECS

[1]Ram Singh, Proceedings of the NCC, Spectral Subtraction Speech Enhancement with RASTA Filtering IIT-B 2007

[2]D P. W. Ellis, PLP, RASTA, MFCC & inversion Matlab, 2005

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

[4]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

[5]Bao Liu, Sherman Riemenschneider, An Adaptive Time- Frequency Representation and Its Fast

Implementation, Department of Mathematics, West Virginia University

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

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

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

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

[10] NitinSawhney, Situational Awareness from Environmental Sounds, SIG, MIT Media Lab, June 13, 1997

[11] C. Valens, IEEE, A Really Friendly Guide to Wavelets, Vol.86, No. 11, Nov 1998,