A ROBUST DEPENDENT AND INDEPENDENT SPEECH RECOGNITION BY USING RASTA-PLP

Dr.V.S.R.Kumari
Principal (Professor)
Affiliated Department of Electronics and Communication Engineering
Sri mittapalli institute of technology for women /JNTUKakinada

Suresh Veesa
Associate Professor
Affiliated Department of Electronics and Communication Engineering
Sri mittapalli institute of technology for women /JNTUKakinada

Srinivasa Rao Chevala
Assistant Professor
Affiliated Department of Electronics and Communication Engineering
Sri mittapalli institute of technology for women /JNTUKakinada

Abstract: A new improved method for speech/speaker recognition is presented in this paper using combination of discrete wavelet transform (DWT) and Relative Spectra Perceptual Linear Prediction (RASTA-PLP) for feature extraction. There are different aspects related to speech like speech recognition, speech verification, speech synthesis, speaker recognition, speaker identification etc. The purpose of this project is to study a speech recognition system using HMM. The objective of the proposed method is to enhance the performance by introducing more features from the signal and applying parallel computations technique leading to an improvement in both the recognition rate and the computational speed whether for a clean or a noisy speech signal for the proposed methods in comparison to using DWT compare with HMM model

Keywords: HMM Model, DWT, median filter RASTA-PLP.

1. INTRODUCTION

The main challenge in a speech/speaker recognition system is to obtain high recognition rate accuracy as possible without taking much time in training and testing the recognition system. The good choice for a feature extraction method, and using the parallel implementation for the neural network classifiers was our way to overcome the above mentioned problems. This was the main idea for our work. The feature extraction process is done by combining the DWT and RASTA-PLP. DWT has been proved that they are well suitable for processing non-stationary signals like speech because of its multi- resolution, multi-scale analysis characteristics. While the main advantage of RASTA-PLP is their robustness to noise. 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

2. EXISTING SYSTEM

We proposed the speech/speaker recognition in this existing method we use the wavelet based HMM model technique. The wave let based is designed in the speech signal. Its base function satisfied s time and bandwidth product least. it was used to make the preprocessing before FFT. in this noise is partially increased.

Anatomical structure of the vocal tract is unique for every person and hence the voice information available in the speech signal can be used to identify the speaker. Recognizing a person by her/his voice is known as speaker recognition. Since differences in the anatomical structure are an intrinsic property of the speaker, voice comes under the category of biometric identity. Using voice for identity

has several advantages. One of the major advantages is remote person authentication. Like any other pattern recognition systems, speaker recognition systems also involve two phases namely, training and testing. Training is the process of familiarizing the system with the voice characteristics of the speakers registering. Testing is the actual recognition task. The block diagram of training phase is shown in Fig.2.1. Feature vectors representing the voice characteristics of the speaker are extracted from the training utterances and are used for building the reference models.

During testing, similar feature vectors are extracted from the test utterance, and the degree of their match with the reference is obtained using some matching technique. The level of match is used to arrive at the decision. The block diagram of the testing phase is given

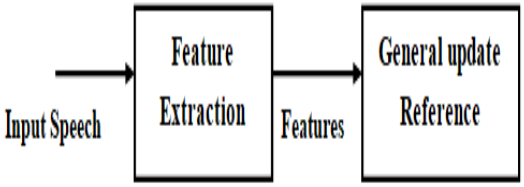

Fig.2.1: *The block diagram of training phase.*

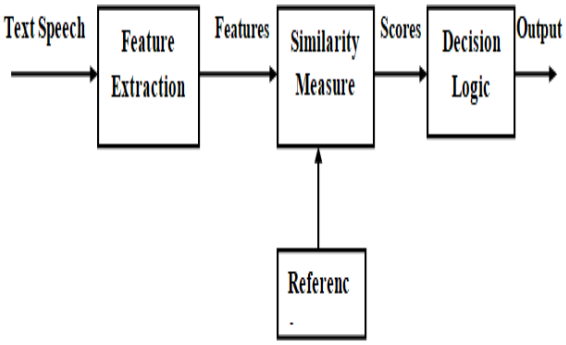

Fig.2.2: *The block diagram of the testing phase*

The extraction and selection of the best parametric representation of acoustic signals is an important task in the design of any speech recognition system; it significantly affects the recognition performance. A compact representation would be provided by a set of mel-frequency cepstrum coefficients (MFCC), which are the results of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale.

3. PROBLEM IDENTIFICATION

- Recognition process is less compare to mimicry voice
- Noise is very high

4. PROPOSED SYSTEM

We proposed the speech/speaker recognition. in this proposed method we use the RASTA-PLP(relative spectral perceptual linear prediction)technique. In this technique we developed by hermansky. While Rasta can overcome both the mg problems resulting from the communication channels by using high pass filter that removes the slowly varying components in each element of the filter bank output

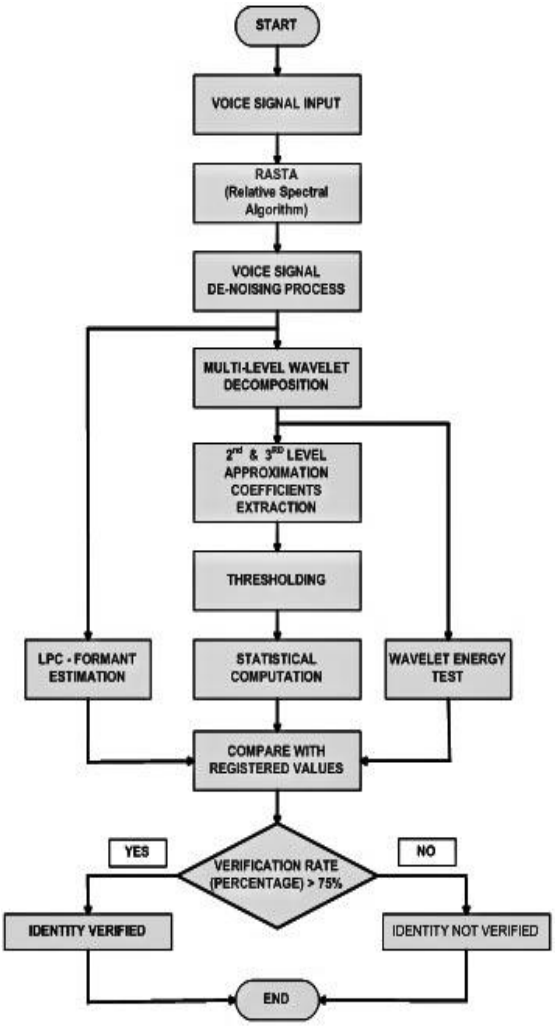

Fig4.1: *System flow chart*

4.1. Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two so called dyadic scales and positions. The mother wavelet is rescaled or dilated by powers of two and translated by integers. Specifically, a function f(t) $\in$ L2(**R**) (defines space of square integrable functions) can be represented as

$$f(t) = \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\, \Psi(2^{-j}t - k) + \sum_{k=-\infty}^{\infty} a(L,K)\emptyset(2^{-L}t - k) \ldots\ldots\ldots\ldots(1)$$

The function ψ(t) is known as the mother wavelet, while φ(t) is known as the scaling

Function. The set of function

$$\{\sqrt{2^{-1}}\,\emptyset(2^{-j}t - l) | j \le L, j, k, L \in Z|\}$$

Where Z is the set of integers is an ortho normal basis for L2(**R**). The numbers a(L, k) are known as the approximation coefficients at scale L, while d(j,k) are known as the detail coefficients at scale j. The approximation and detail coefficients can be expressed as:

$$d(L,K) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\, \emptyset(2^{-L}t - k)dt \ldots.(2)$$

$$d(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\, \Psi(2^{-j}t - k)dt \ldots(3)$$

To provide some understanding of the above coefficients consider a projection fl(t) of the function f(t) that provides the best approximation (in the sense of minimum error energy) to f(t) at a scale l. This projection can be constructed from the coefficients a(L, k), using the equation

$$f_1(t) = \sum_{k=-\infty}^{\infty} a(L,K)\emptyset(2^{-L}t - k) \ldots\ldots(4)$$

As the scale l decreases, the approximation becomes finer, converging to f(t) as l → 0.□ The difference between the approximation at scale l + 1 and that at l, fl+1(t) - fl(t), is completely described by the coefficients d(j, k) using the equation

$$f_{l+1}(t) - f_1(t) = \sum_{k=-\infty}^{\infty} d(l,k)\varphi(2^{-l}t - k) \ldots(5)$$

Using these relations, given a(L, k) and {d(j, k) | j ≤ L}, it is clear that we can build the approximation at any scale. Hence, the wavelet transform breaks the signal up into a coarse approximation fL(t) (given a(L, k)) and a number of layers of detail {fj+1(t)-fj(t)| j < L} (given by {d(j, k) | j ≤ L}). As each layer

of detail is added, the approximation at the next finer scale is achieved.

4.2. Vanishing Moments

The number of vanishing moments of a wavelet indicates the smoothness of the wavelet function as well as the flatness of the frequency response of the wavelet filters (filters used to compute the DWT).Typically a wavelet with p vanishing moments satisfies the following equation .

$$\int_{-\infty}^{\infty} t^m \varphi(t)dt = 0 \qquad for\ = 0, \ldots\ldots, p-1$$

Or equivalently,

$$\sum_k (-1)^k k^m c(k) = 0\ for\ \ m = 0, \ldots . p-1$$

For the representation of smooth signals, a higher number of vanishing moments leads to a faster decay rate of wavelet coefficients. Thus, wavelets with a high number of vanishing moments lead to a more compact signal representation and are hence useful in coding applications. However, in general, the length of the filters increases with the number of vanishing moments and the complexity of computing the DWT coefficients increases with the size of the wavelet filters. These are the approximation coefficients cA1 (low frequency information) and the detail coefficients cD1 (high frequency information), as shown in the figure below.

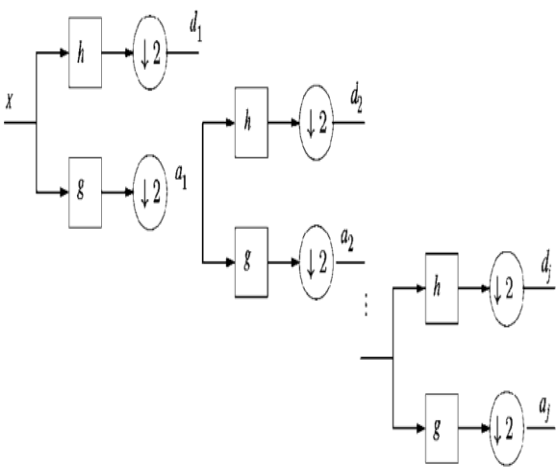

Fig.4.2: *Filtering operation of the DWT*

4.3. Implementation Using Filters

To explain the implementation of the Fast Wavelet Transform algorithm consider the following equations:

$$\emptyset(t) = \sum_k c(k)\varphi(2t-k) \ldots\ldots\ldots (6)$$

$$\varphi(t) = \sum_k (-1)^k c(1-k)\emptyset(2t-k) \ldots . (7)$$

$$\sum_K C_K C_{K-2M} = 2\delta_{0,m} \ \ldots\ldots \ldots\ldots . (8)$$

The first equation is known as the twin-scale relation (or the dilation equation) and defines the scaling function φ. The next equation expresses the wavelet ψ in terms of the scaling function φ. The third equation is the condition required for the wavelet to be Orthogonal to the scaling function and it's translating function.

4.4. Multilevel Decomposition

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution Components. This is called the wavelet decomposition tree.

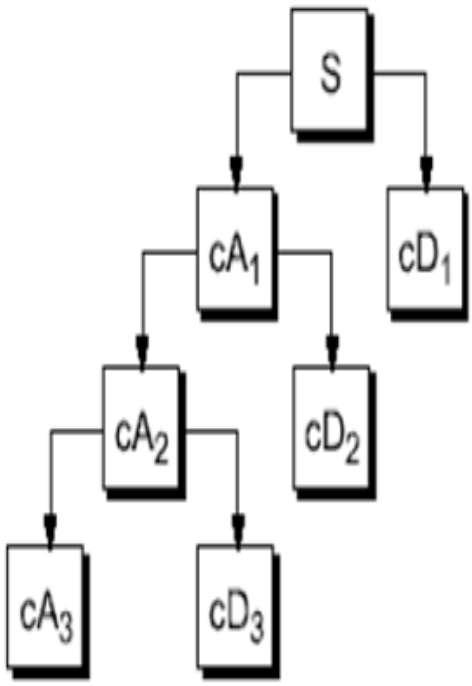

Fig4.3: *Decomposition of DWT co-efficient*

The wavelet decomposition of the signals analyzed at level j has the following structure [C_{Aj}, C_{Dj}, ..., C_{D1}]. Looking at a signals wavelet decomposition tree can reveal valuable information.

4.5 Signal Reconstruction

The original signal can be reconstructed or synthesized using the inverse discrete wavelet transform (IDWT). The synthesis starts with the approximation and detail coefficients c_{Aj} and C_{Dj}, and then reconstructs C_{Aj-1} by up sampling and filtering with the reconstruction filters.

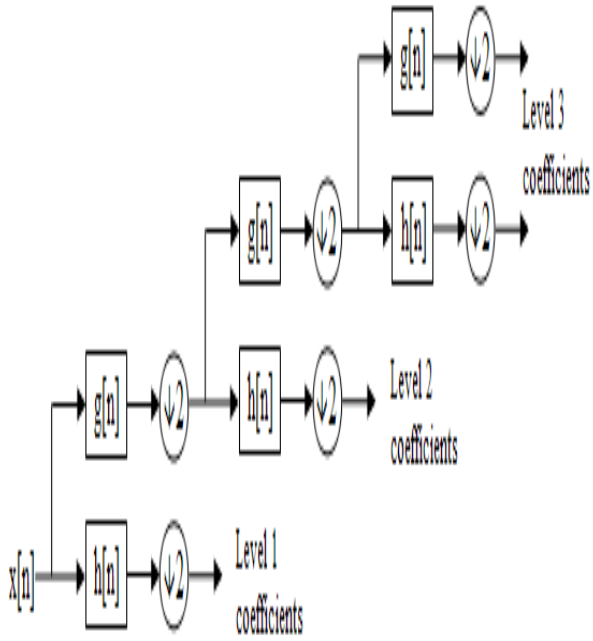

Fig 4.4: *Filter levels*

The reconstruction filters are designed in such a way to cancel out the effects of aliasing introduced in the wavelet decomposition phase. The reconstruction filters (Lo_ R and Hi_ R) together with the low and high pass decomposition filters, forms a system known as quadrature mirror filters (QMF).

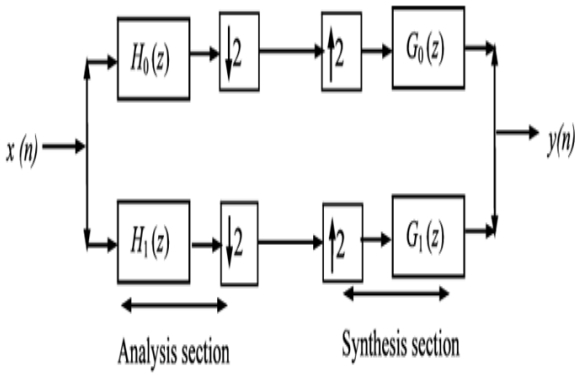

Fig4.5: *Quadrature mirror filters (QMF)*

4.6. Feature extraction: 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. 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.

5. RESULTS

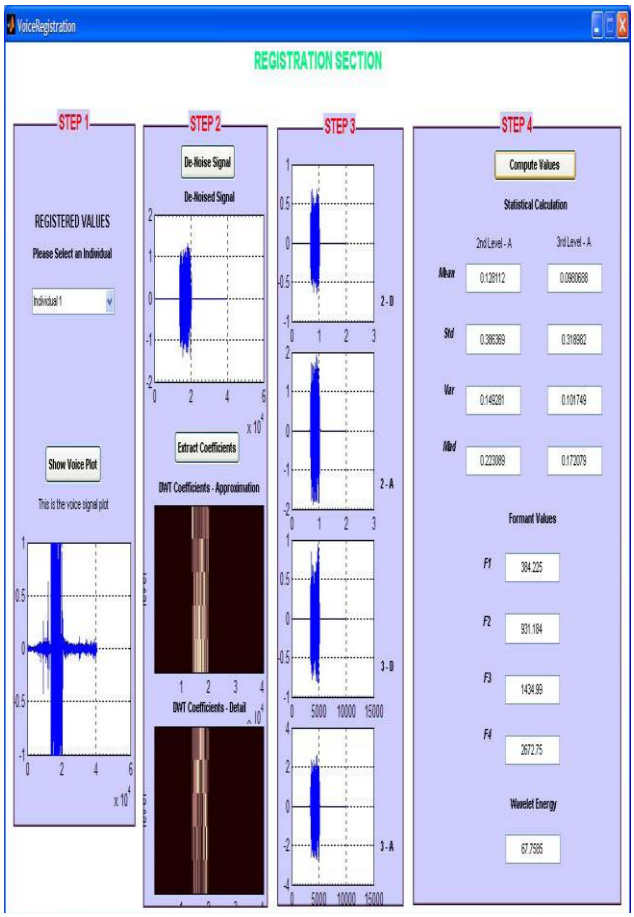

Fig5.1: Voice Registration GUI

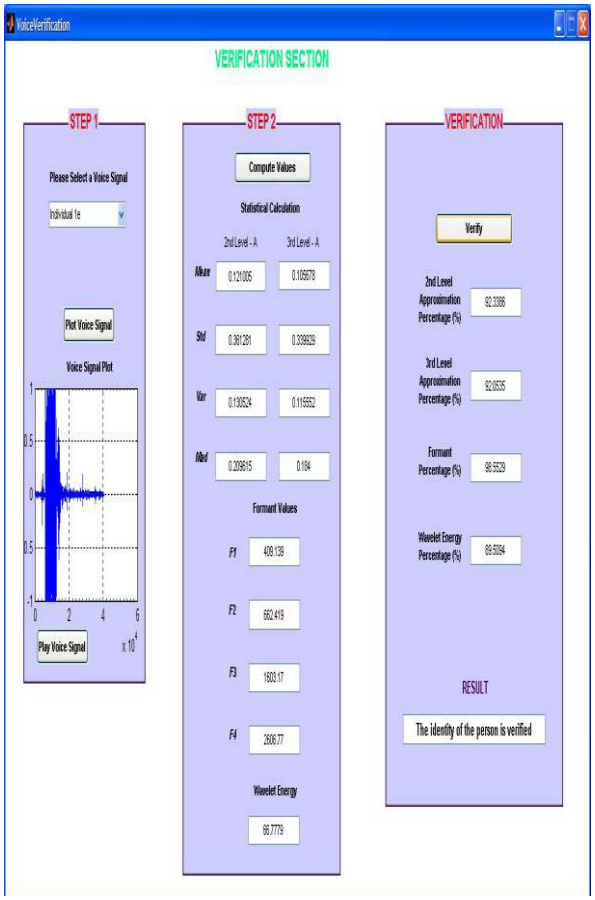

Fig 5.2: Voice Verification GUI

Applications

1) Education System

2) Industrial

3) ATM and Banking

4) Electronic (Laptop, Mobile)

6. CONCLUSION

The speech signal pre-processing is carried out by different methods. Paper shows that, the feature extracted with wavelet transforms has the highest accuracy for recognition of words. This has been proved in both the conditions for clean and noisy speech signals. It is also shown that the RASTA-PLP method used to verify dependent and independent using Wavelets for noisy speech yields better results. the verification tests have been carried and an accuracy rate of approximately 90 % has been achieved by proposed algorithm It is clearly compare to HMM model

REFERENCES

[1]. SD.B.Paul,"Speech Recognition usingHidden Markov Model. International Magazine of Computers and Technology, No. 8, 2014 Vol. 13

[2]. M.A.Anusuya , S.K.Katti "Speech Recognition by Machine: A Review" International journal of computer science and Information Security 2009

[3]. L.R.Rabiner and B.H.jaung ," Fundamentles of Speech Recognition Prentice-Hall, Englewood Cliff, New Jersy,1993.

[4]. Mahdi Shaneh, and Azizollah Taheri,"Voice Command Recognition System Based on MFCC and VQ Algorithms", World Academy of Science, Engineering and Technology 57 2009

[5] H. Combrinck and E. Botha, "On the mel-scaled cepstrum," Video Communications, SPIE Electronic Imaging, UCRL Conf. San Jose, vol.200706, Jan 2004.

[6] Electronic Engineering, University of Pretoria.,Journal of Computer Science 3 (8): 608-616, 2007 ISSN 1549-3636.

[7] Ahmad A. M. Abushariah,Teddy S. Gunawan, Othman O. Khalifa"English Digits Speech Recognition System Based on Hidden Markov Models", International Islamic University Malaysia, International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[8] AnjaliBala,ABHIJEETKumar,Nidhika Birla,"Voice command recognition System Based on MFCC and DTW",International Journal of Engineering Science and Technology,Vol.2(12),2010

[9] Ibrahim Patel,Dr.Y.Srinivasa Rao, , "Speech recognition using Hidden Markov Model With MFCCSubband Technique." 2010 International Conference on Recent Trends in Information,Telecommunication and Computing.

[10] ShumailaIqbal,Tahira,Mehboob,Malik ,"Voice Recognition using HMM with MFCC for secure ATM",IJCS Vol.8,Issue 6 Nov 2011

[11] Vimala C, Dr.V.Radha, "A Review on Speech Recognition Challenges and Approaches", World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 1, 1-7, 2012

[12] Lawrence R. Rabiner, Fellow, IEEE 'A Tutorial On Hidden Markov Model And Selected Applications In Speech Recognition, Proceedings Of The IEEE, Vol. 77, No. 2, February 1989.