International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Voice Recognition using Improved Relative Spectral Algorithm

N Nagaraju[1], L Shruthi[2]

[1]*Assisstant Professor, Dept. of Electronics and Communication Engineering, Institute of Aeronautical Engineering, Hyderabad, Telangana, India*
[2]*Assisstant Professor, Dept. of Electronics and Communication Engineering, Institute of Aeronautical Engineering, Hyderabad, Telangana, India*

Abstract - *For further proficient depiction of the speech signal, the relevance of the wavelet analysis is considered. This survey presents a valuable and strong technique for extracting features for speech processing. At this point, we projected a new human voice identification scheme with the amalgamation of Discrete Wavelet (DW) and Relative Spectra Algorithm with Linear Predictive coding. Initially, we will examine the proposed techniques to exercise speech signals and then outline a train characteristic vector which contains the low level features extracted, wavelet and linear predictive coefficients. Afterwards, the identical method will be applied to the testing speech signals and resolve to figure a test characteristic vector. At present, we will compare the two characteristic vectors by calculating the Euclidean distance among the vectors to recognize the speech and speaker. If the distance among two vectors is close to zero then the tested speech/speaker will be in line with the educated speech/speaker. Simulation results have been compared with LPC method, and shown that the anticipated proposal has performed better-quality to the presented system by using the fifty preloaded voice signals from four folks, the authentication tests have been conceded and an exactness rate of just about 90 % has been achieved.*

Key Words: **Voice Recognition, DW, LPC, RASTA, Euclidean Distance**

1. INTRODUCTION

The acoustic signal particularly tone of voice signal is suitable one of the key component in human's everyday life. The fundamental purpose of the voice signal is that it is used as one of the most important tools for communication. On the other hand, owing to scientific development, the voice signal is more processed by means of software applications in addition to the voice signal information is utilized in a variety of applications. The elementary proposal of this development is to make use of wavelets as a mean of extracting features from a voice signal. The wavelet procedure is well thought-out a moderately innovative procedure in the field of signal processing compared to further methods or techniques at present working in this field. Existing methods used in the field of signal processing consist of Fourier Transform and Short Term Fourier Transform (STFT) [1] [2]. On the other hand owing to strict boundaries forced by both the Fourier Transform and Short Term Fourier Transform in analyzing signals deems them hopeless in analyzing composite and self-motivated signals such as the voice signal [3][4]. With the intention of replacement the shortcomings forced by both the ordinary signal processing methods, the wavelet signal processing procedure is used. The wavelet procedure is used to pull out the features in the voice signal by dealing out information at diverse scales. The wavelet procedure manipulates the scales to provide a superior connection in detecting a variety of frequency components in the signal. These features are after that more processed with the intention to create the voice identification scheme. Speech identification is the development of robotically extracting and determining linguistic information conveyed by a speech signal by means of computers or electronic circuits.

2. SPEECH RECOGNITION

Speech identification systems can be secret according to the subsequent categories:

2.1 Speaker Dependent vs. Speaker Independent

A speaker-dependent speech detection system is single so as to guide to be acquainted with the verbal communication of no more than one speaker. Such systems are convention built in support of just a particular human being, in addition to be not commercially feasible. On the contrary, a speaker-independent arrangement is one that self-determination is rigid to accomplish, as speech identification systems are inclined to be converted into adjusted in the direction of the speakers they are skilled on, ensuing in fault rates that are superior to speaker dependent systems.

2.2 Isolated vs. Continuous

In isolated speech, the speaker pauses temporarily flanked by each utterance, at the same time in continuous speech the speaker speaks in uninterrupted in addition to maybe extended stream, by means of small or no breaks in between. Isolated speech identification systems are straightforward to fabricate, as it is insignificant to decide where one utterance ends and one more starts, in addition to every word tends to be extra cleanly and obviously

spoken. Vocabulary spoken in uninterrupted speech on the other hand is subjected to the co-articulation consequence, during which the articulation of a utterance is customized by the vocabulary adjacent it. This makes exercise a speech scheme not easy, as there might be several incompatible pronunciations for the identical utterance.

3. WAVELET TRANSFORM

Wavelets are arithmetical functions so as to gratify certain necessities [5]. Since a arithmetical opinion, the wavelet is described as a task so as to incorporate to zero plus it has a waveform that has a inadequate interval. The wavelet is in addition to predetermined length which means that it is efficiently supported. Wavelets examine a signal using different scales. This move towards signal processing is called multi-resolution STFT. On the other hand the signal is not segmented or alienated uniformly by using a predetermined window span. Multi-Resolution Analysis (MRA) examine the frequency mechanism of the signal with dissimilar resolutions. This move towards particularly makes sense for non-periodic signal such as the voice signal which has low-frequency components dominating for extensive durations in addition to short durations of high-frequency components [6]. A great scale can be interpreted as a "huge" window. By means of a large scale to examine the signal, the disgusting features of a signal can be obtained. Vice versa, a small scale is interpreted as a "thin" window and the small scale is used to notice the superior information of the signal. This property of wavelet examination makes it very dominant and helpful in detecting or enlightening unknown aspects of information and in view of the fact that wavelet transform provides a different point of view in analyzing a signal, compression or de-noising a signal can be conceded out exclusive to a great extent signal deprivation. Local features of a signal can be detected with remote enhanced accurateness with wavelet transform [7].

3.1 Fourier Transform

The signal can be analyzed further efficiently in frequency domain than the point in time area, for the reason that the uniqueness of a signal will be supplementary in frequency domain. One feasible technique to translate or alter the signal from time to frequency domain is Fourier transform (FT). FT is an approach which breaks down the signal into different frequencies of sinusoids and it is distinct as a arithmetical move towards for transforming the signal from time domain to frequency domain. FT has a disadvantage that it will exercise for only at standstill signals, which will not be dissimilar with the time period. for the reason that, the FT useful for the complete signal except not segments of a signal, if we think about non-stationary signal the signal will vary with the time period, which might not be altered by FT, and one more disadvantage that we comprise with the FT.

3.2 Short-Time Fourier Transform

On the way to correct the shortage in FT, Dennis Gabor in 1946 introduced a new method called windowing, which can be applied to the signal to investigate a tiny segment of a signal. This version has been called as the Short-Time Fourier Transform (STFT), in which the signal will be mapped into time and frequency information. In STFT, the window is predetermined. So, this window will not change with the time period of the signal i.e., for both fine resolution and extensive resolution. In addition to cannot envisage the frequency substance at each time interval division. On the way to triumph over the drawbacks of STFT, a wavelet method has been introduced by way of changeable window size. Wavelet analysis allows the use of long time intervals where additional accurate low-frequency information, and shorter regions where we want high-frequency information. In figure1 it is shown that the contrast of FT, STFT and wavelet transform by taking into account an instance input signal and how the examination of alteration techniques resolve apply to get the frequency information of input signal. Observe that in wavelet analysis the graphical demonstration shows that the wavelet has additional number of features than the FT and STFT. Wavelet is also called as multi resolution analysis (MRA). Here are this looks like in distinction with the time-based, frequency-based, and STFT views of a signal.

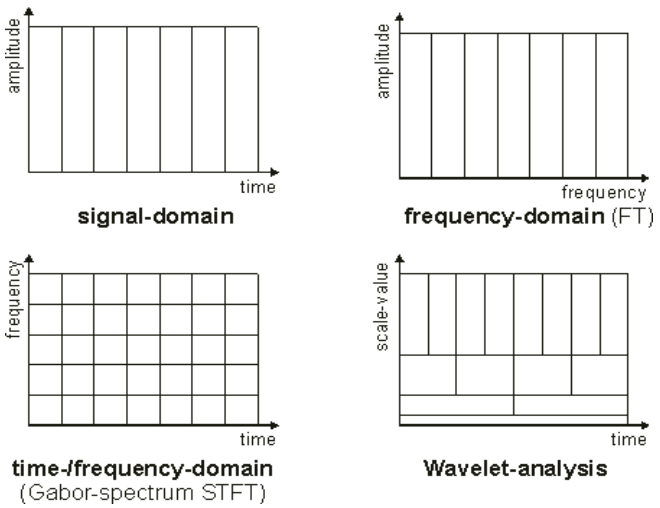


Fig -1: Comparison of FT, STFT and Wavelet Analysis of a Signal

3.3 Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is a revised edition of Continuous Wavelet Transform (CWT). The DWT compensates for the enormous quantity of information generated by the CWT. The fundamental operation of DWT is similar to the CWT on the other hand the scales used by the wavelet and their positions are based upon powers of two.

At the same time as in numerous real world applications, the majority of the essential features of a signal stretch out in the low frequency sector. For voice signals, the low frequency substance is the segment or the part of the signal that gives the signal its distinctiveness whereas the high frequency substance can be considered as the fraction of the signal that gives fine distinction to the signal. This is analogous to imparting essence to the signal. For a voice signal, if the high frequency content is detached, the voice will sound unusual but the significance can still be heard or conveyed. This is not accurate if the low frequency content of the signal is detached as what is being vocal cannot be heard apart from only for some random noise. The DWT is represented using the algebraic equation

$$W(\tau,s) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} \Psi\left(\frac{t-\tau}{s}\right) dt \text{ — (1)}$$

$$\int_{-\infty}^{\infty} \Psi(t)dt = 0 \text{ — (2)}$$

$$\int_{-\infty}^{\infty} \left|\left(\Psi(t)\right)^2\right| dt < \infty \text{ — (3)}$$

The fundamental process of the DWT is that the signal is conceded from beginning to end a sequence 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 [10]. This means the approximations are obtained by means of the high scale wavelets which communicate to the low frequency. The high frequency components of the signal called the fine points are obtained by using the low scale wavelets which corresponds to the high frequency. From figure2, demonstrates the single level filtering using DW. Primary the signal is fed into the wavelet filters, these wavelet filters comprises of both the high-pass and low-pass filter. After that, these filters will break up the high frequency content and low frequency content of the signal. On the other hand, with DW the numbers of samples are condensed according to 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 dispensation power [11] the DW is chosen due its simplicity and ease of operation in treatment composite signals such as the voice signal.

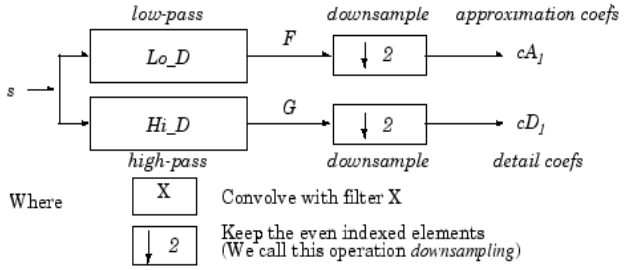


Fig -2: Demonstration of single level filtering using Discrete Wavelet

3.4 Wavelet Energy

At whatever time a signal is being festering by means of the wavelet disintegration technique, at hand is a definite quantity or fraction of energy being retained by mutually the estimate and the feature. This energy can be obtained from the wavelet accounting vector and the wavelet disintegration vector. The energy calculated is a ratio as it compares the original signal and the decomposed signal.

4. EXISTING ALGORITHM

The LPC (Linear Predictive Coding) technique is derived from the word linear prediction. Linear prediction as the word implies is a type of arithmetical process. This arithmetical function which is used in discrete time signal estimates the expectations principles based upon a linear function of preceding samples [8].

$$\hat{x}(n) = -\sum_{i=1}^{p} a(i)x(n-i) - \text{ (4)}$$

$\hat{x}(n)$ is the predicted or estimated value and $x(n-i)$ is the previous value. By expanding this equation

$$\hat{x}(n) = -a_1 x(n-1) - a_2 x(n-2) - a_3 x(n-3) \ldots \text{ ----- (5)}$$

The LPC will examine the signal by estimating or predicting the formants. Subsequently, the formants effects are detached from the speech signal. The strength and frequency of the residual buzz is predictable. By removing the formants from the voices signal will enable us to get rid of the reverberation 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 required. The coefficients are predictable by taking the mean square error between the predicted signal and the original signal. By minimizing the error, the coefficients are detected with a superior accurateness and the formants of the voice signal are obtained.

5. VOICE SIGNAL ANALYSIS

5.1 RASTA (Relative Spectral Algorithm)

RASTA or Relative Spectral Algorithm is recognized as a procedure that is developed as the early phase for voice identification [13]. This technique works through applying a band-pass filter to the energy in every frequency sub-band in order to smooth over short-term noise variations and to eliminate whichever invariable offset. In voice signals, immobile noises are regularly detected. A stationary noise that are in attendance for the full period of a positive signal and does not have withdrawing feature [14]. Their possessions does not transform in excess of time. The postulation that desires to be finished is that the noise varies gradually with respect to speech. This makes the RASTA an ideal tool to be incorporated in the early

stages of voice signal filtering to eliminate stationary noises [15].

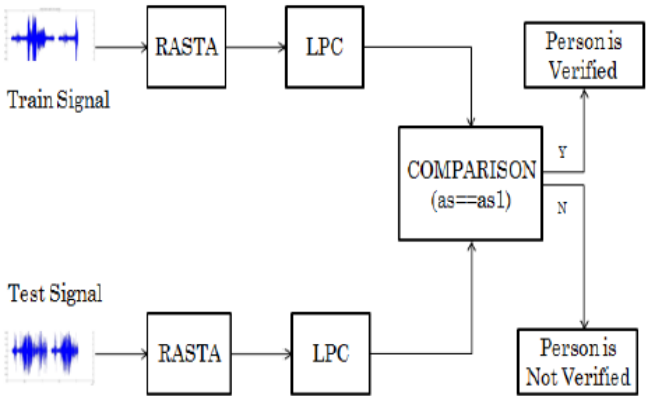


Fig -3: Block Diagram of LPC based recognition system

The stationary noises that are recognized are noises in the frequency range of 1Hz - 100Hz.

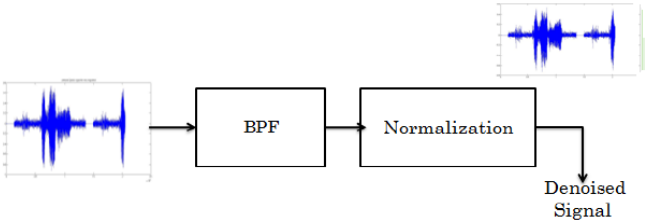


Fig -4: Block diagram of RASTA process

5.2 Formant Estimation

Formant is one of the major components of speech. The frequencies at which the reverberating peaks occur are called the formant frequencies or simply formants [12]. The formant of the signal can be obtained by analyzing the vocal territory frequency response. Figure 5 shows the vocal territory 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 arithmetical operation.

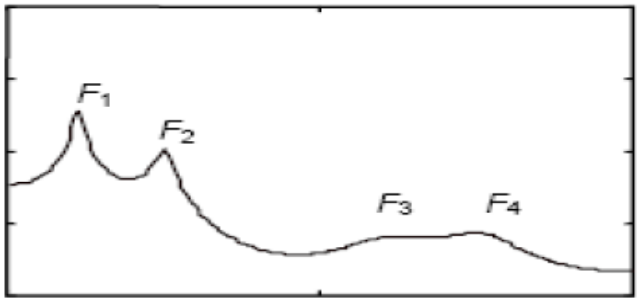


Fig -5: Vocal tract frequency response

5.3 RASTA-LPC and DWT Implementation

With the intention of realize the structure; a certain tactic is implemented by decomposing the voice signal to its estimate and feature. On or after the estimate and feature coefficients that are extracted, the tactic is implemented in order to bring out the identification process. The planned line of attack for the identification phase is the geometric computation. Four different types of statistical calculations are conceded 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 indomitable through several trial and errors. The coefficients that are extracted from the wavelet disintegration method is the second level coefficients as the level two coefficients enclose most of the interrelated data of the voice signal. The data at higher levels contains very little quantity of data deeming it impracticable for the acknowledgment phase. For this reason initial system accomplishment, the level two coefficients are used.

The coefficients are additional entry to take away the low correspondence values, and using this coefficients statistical calculation 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 fraction of confirmation is premeditated by comparing the existing signal values against the registered voice signal values. The percentage of verification is given by:

Verification % = (Test value / Registered value) x100 -- (6)

Between the tested and registered value, whichever value is superior it is taken as the denominator and the lesser value is taken as the numerator. Figure 6 shows the complete flowchart which includes all the significant system components that are used in the voice authentication program.

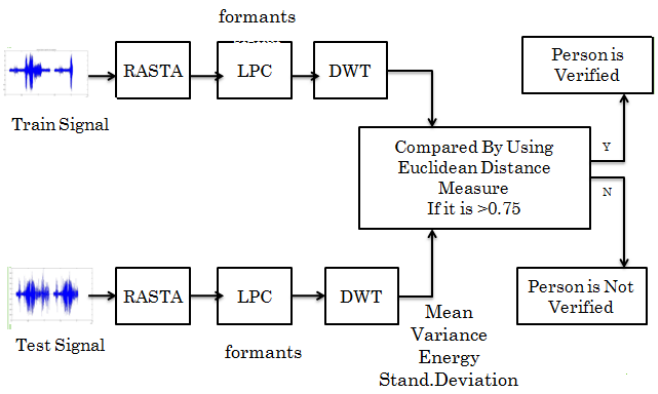


Fig -6: Block diagram of proposed text dependent speaker identification system

6. SIMULATION RESULTS

Experimental results have been shown for various voice test signals with LPC and proposed algorithms. All the experiments have been done in MATLAB 2011a version with 4GB RAM and i5 processor for speed specifications.

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad — (7)$$

$$\text{Standard Deviation} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$$

$$\text{Variance} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \quad — (9)$$

Table -1: Comparison test

Individual	1	2	3	4
1	V	NV	NV	NV
2	NV	V	NV	NV
3	NV	NV	NV	NV
4	NV	NV	NV	V

V= Verified NV= Not Verified

6.1 Test signal-1

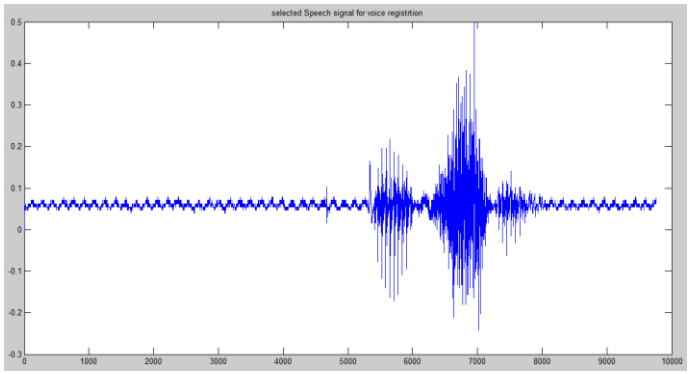


Fig -7: Original speech signal for voice registration

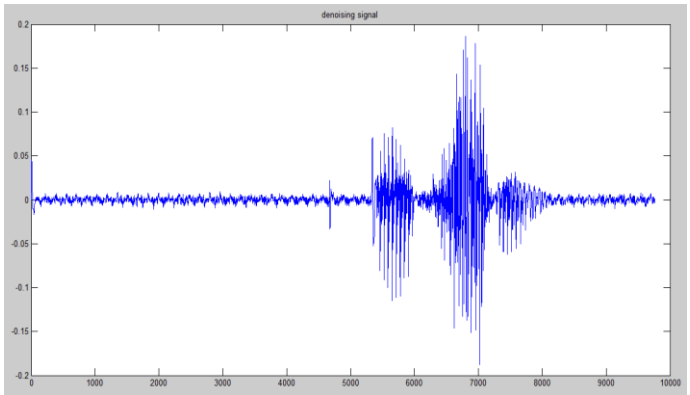


Fig -8: Denoised signal for voice registration

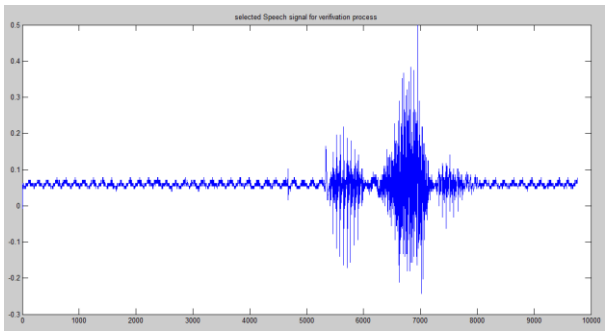
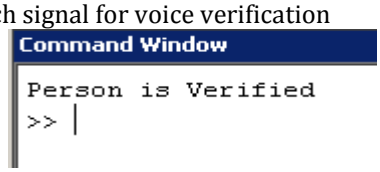


Fig -9: Speech signal for voice verification



6.2 Test signal-2

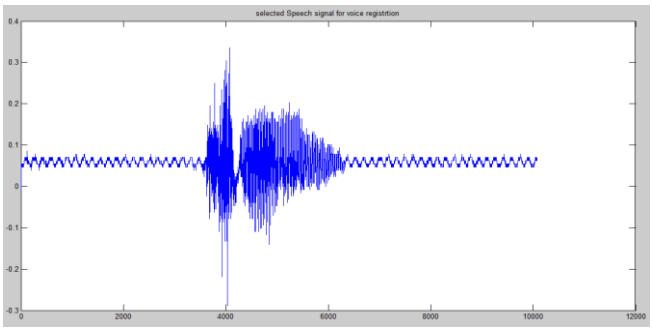


Fig -10: Original speech signal for voice registration

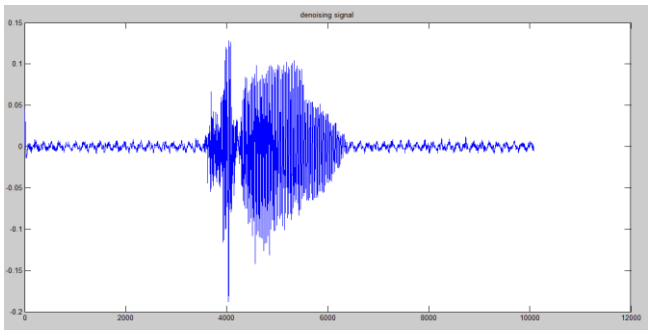


Fig -11: Denoised signal for voice registration

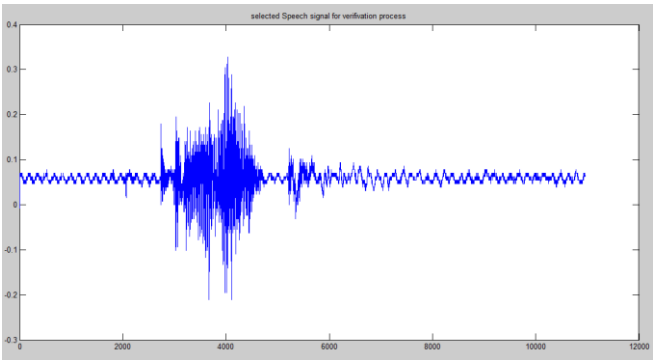
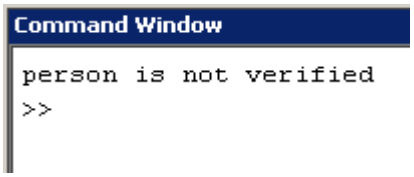


Fig -12: Speech signal for voice verification



From the Table above of the verification result shows from the five random tests carried out, at any one given time, the program can successfully verify 3 out 4 persons accurately. The complete systems which constitutes all the system components for the recognition methodology is one of the main reasons for the high accuracy of the system. Currently, the percentage of verification is set at an average value of 75%. The verification rate can be further increased or decreased by adjusting the percentage of verification to a higher or lower value. By substituting a lower value, the system will be less secure while a higher value could jeopardize the accessibility rate of the system because of the certain level of tolerance is required for the voice signal as it tends to change with internal and external factors.

7. CONCLUSION

The Voice 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 dependant 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. By using the fifty preloaded voice signals from six individuals, the verification tests have been carried and an accuracy rate of approximately 90 % has been achieved by proposed algorithm. The system can be enhanced further by using advanced pattern recognition techniques such as Neural Network or Hidden Markov Model (HMM) for more accurate and efficient system.

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 2011.

[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] Pala Mahesh Kumar, A New Human Voice Recognition System, AJSAT, July, 2016.

[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, University Of Sannio, Italy.

[9] Michael Unser, Thierry Blu, IEEE Transactions on Signal Processing, Wavelet Theory Demystified, Vol. 51,No. 2,Feb'13.

[10] C. Valens, IEEE, A Really Friendly Guide to Wavelets, Vol.86, No. 11, Nov 2012.

[11] James M. Lewis, C. S Burrus, Approximate CWT with An Application To Noise Reduction, Rice University, Houston.

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

[13] Ram Singh, Proceedings of the NCC, Spectral Subtraction Speech Enhancement with RASTA Filtering IIT-B 2012.

[14] NitinSawhney, Situational Awareness from Environmental Sounds, SIG, MIT Media Lab, June 13, 2013.

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