ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

A NOVEL APPROACH FOR HYBRID WAVELET FOURIER-HMM SPEAKER RECOGNITION VOICE VERIFICATION SYSTEM USING WAVELETS

[1] N.Pooja, [2] S.V.S.Jaya Shyam

[1]PG Student, Department of ECE, Sri Venkateswara College of Engineering for Women, pooja.nagirimadugu66@gmail.com , Tirupati, India.

[2] Assistant Professor, Department of ECE, Sri Venkateswara College of Engineering for Women, svsjsham@gmail.com , Tirupati, India

Abstract -The paper presents successful experiments on combining two speaker recognition methods into a hybrid system. The rest branch of recognition is an innovative approach based on discrete wavelet-transform. The second one is classic, based on HTK and classication into voice and unvoiced segments. The hybrid solution outperforms both on a small test set. Further tests of this solution will be conducted and reported.. Voice is compared four out of five individual are verified and the result show higher percentage of accuracy. The accuracy of the system can be increased by using a advance pattern recognition The latest technology improvements in such elds as banking, communications and networking require the latest advances in security systems. In the last years a new kind of biometric identication has risen among the others as it is printed in subjects' bodies, it is impossible to lose, almost impossible to duplicate and univocally designates a person. The biometric keys exploit some human characteristics that are dierent and unique in each person such as fingerprints, DNA chains, face shape and voice. Technique such as Hidden Markov Model (HMM).

Keywords - STFT, LPC, RASTA, Voice Verification model

INTRODUCTION

The latest technology improvements in such elds as banking, communications and networking require the latest advances in security systems. In the last years a new kind of biometric identication has risen among the others as it is printed in subjects' bodies, it is impossible to lose, almost impossible to duplicate and univocally designates a person. The biometric keys exploit some human characteristics that are different and unique in each person such as fingerprints, DNA chains, face shape and voice. Voice is a phenomenon that is highly dependent on the speaker. Many physical aspects of speech such as the timber, tone or intensity vary a lot from a speaker to another one. The same happens with other linguistic aspects as the intonation and range of vocabulary or expressions a speaker normally uses. All these properties make voice a very powerful biometric key to be used in security systems since the physical characteristics of speech are easy to measure in comparison to other biometric keys. In addition, the speech signal has been deeply studied for many years so many powerful algorithms are found to deal with this kind of signal. Signal deems them ineffective in analyzing complex and dynamic signal such as the voice signal [3] [4]. To substitute the short comings of the common signal processing method, wavelet signal technique is used. Wavelet technique is used to extract the features of the voice signal by processing it in different scales.

The wavelet technique designs the scales to given a higher correlation in detecting the various frequency components in the voice signal. These voice signals are processed in order to construct the voice recognition system. Extracting features in the voice signal provides a open door for the different application benefit from the voice extracted feature. Application such as speech to text translators speech recognition and voice based security system are some of the feature systems that can be developed.

WAVELET TRANSFORM

A good biometric key has to match certain requirements. It has to be easy to extract, measure, save and compare. Voiceprints match all these requirements since not a very

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

expensive hardware is needed to perform all these operations. In fact, the only infrastructure needed is a microphone and a PC. A medium-quality microphone is relatively cheap hardware if one compares it to a digital camera, iris ornery scanner, not mentioning a DNA analyzer. In addition, many new banking applications rely on the usage of telephone line. Voiceprints are the only suitable biometric technology able to operate in this environment. There are a number of features in a speech signal which make recognizing a speaker possible. The appropriate transformation of speech is an important problem because the representation in the time domain gives little information about the speech signal properties. It is necessary to obtain the optimal spectral representation. Usually, methods which are based on the Fourier or wavelet transform are used. In this way, the frequency properties of speech are analyzed. To improve recognition, we tested a hybrid system based on Wavelet-Fourier Transform (WFT) and more traditional HTK based system on Mel Frequency Cepstral Coecients (MFCC) with Wavelets are mathematical functions that satisfy certain necessity. From mathematical point of view, the wavelet is a function that should integrate at zero and it has a waveform that has a limited duration. The wavelet has finite length which means that it is compactly supported. It analyzes the signal using different scales. This approach in signal processing is called as multi resolution analysis. The scale is similar to the window function in STFT. The signal is not divided or segmented using fixed window size. In Multi-Resolution Analysis (MRA) analyze the resolution with different frequency component of the signal. This approaches mainly sense for non-periodic signal such as voice signal which has high frequency components dominates for short duration and low frequency components dominates segmented using fixed window size. In Multi-Resolution Analysis (MRA) analyze the resolution with different frequency component of the signal. These approaches for long duration [6]. Small scales can be interpreted as "narrow" window. Using small scale fine details of the signal is analyzed. Vice versa, A large scales signal is used for "large window". Using large scale gross feature of the signal is analyzed. This property of the wavelet makes it more powerful and useful in detecting revealing hidden aspects of the data. Wavelet transform provides different aspects in analyzing a signal, compression or de-noising a signal can be carried out without much signal degradation. Local feature of the signal can be detected with higher accuracy in wavelet transform. Frequency content is removed then the voice will sound differently but the message can be heard. This is not true that if low frequency component is removed than what is being is spoken cannot be heard expect only for some random noise. The DWT can be given using the mathematical equation

$$\psi_{L,k}(n) = 2^{-\frac{1}{2}}\psi(2^{-1}n - k) \quad (2)$$

$$C(a,b) = C(j,k) = \sum_{n-1}^{\cdot} f(n)\psi_{j,k,}(n) \; (1)$$

The basic function of the DWT is to pass the signal through the series high pass and low pass filter to obtain the high frequency and low frequency content of the signal. The low frequency of the signal is called as the approximations .This means that the approximations are obtained by the high scales wavelet which corresponds to the low frequency. The high frequency components are called as the details. This is obtained by using the low scale wavelet which corresponds to the high frequency.

From Figure 1.Shows the single level filtering using DWT. First the signal is employed into the wavelet filters. This wavelet provides the comparison of the both low pass filter and the high pass filter. Then these filters will separate the high frequency component and low frequency component present in the signal. In DWT the number of the sample are reduced according to the dyadic scale. This process is called as sub-sampling. Sub-sampling means reducing the samples by a given factor. Due to the disadvantage in the CWT it requires more processing power then the DWT. DWT is selected for the simplicity and ease of operation in handling complex signal such as voice signal

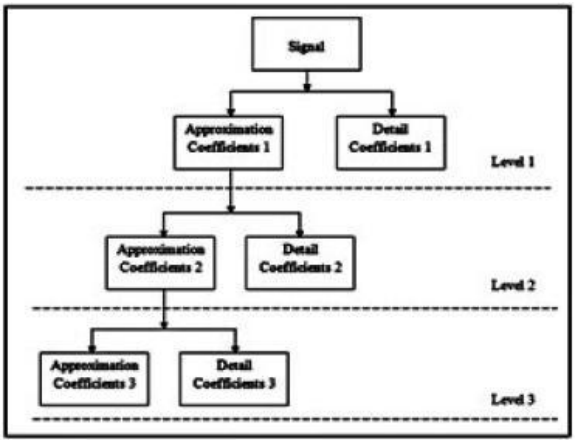


Fig. 1 DWT operation

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

B. Wavelet Energy

Signal is decomposed using the wavelet decomposition method; there are certain amount of percentage of energy that is being retained by both the approximation and the detail. This energy can be obtained by the wavelet decomposition vector and book keeping vector. The energy calculated as the ratio of the compared original signal and the decomposed signal.

VOICE SIGNAL ANALYSIS

Voice problems that require voice analysis most commonly originate from the vocal folds or the laryngeal musculature that controls them, since the folds are subject to collision forces with each vibratory cycle and to drying from the air being forced through the small gap between them, and the laryngeal musculature is intensely active during speech or singing and is subject to tiring. However, dynamic analysis of the vocal folds and their movement is physically difficult. The location of the vocal folds effectively prohibits direct, invasive measurement of movement. Less invasive imaging methods such as x-rays or ultrasounds do not work because the vocal cords are surrounded by cartilage which distorts image quality. Movements in the vocal cords are rapid, fundamental frequencies are usually between 80 and 300 Hz, thus preventing usage of ordinary video.

A. RASTA (Relative Spectral Algorithm)

\

Relative Spectral Algorithm or RASTA is a technique to develop the initial stage for voice recognition [13]. This method works by applying the band pass filter to the energy in each frequency sub-band in order to a smooth over short-term noise variation and to remove any constant offset. In Stationary signal often stationary noise is detected. Stationary noise is present in the full period of the certain signal which does not have diminishing feature[14]. Their property does not change over time. An assumption is to be made that the noise varies slowly with respect to the speech. This makes the RASTA algorithm to be employed in the earlier stages of the voice signal to filter the stationary noises [15]. The stationary b\noises will be in the range of the 1Hz – 100Hz

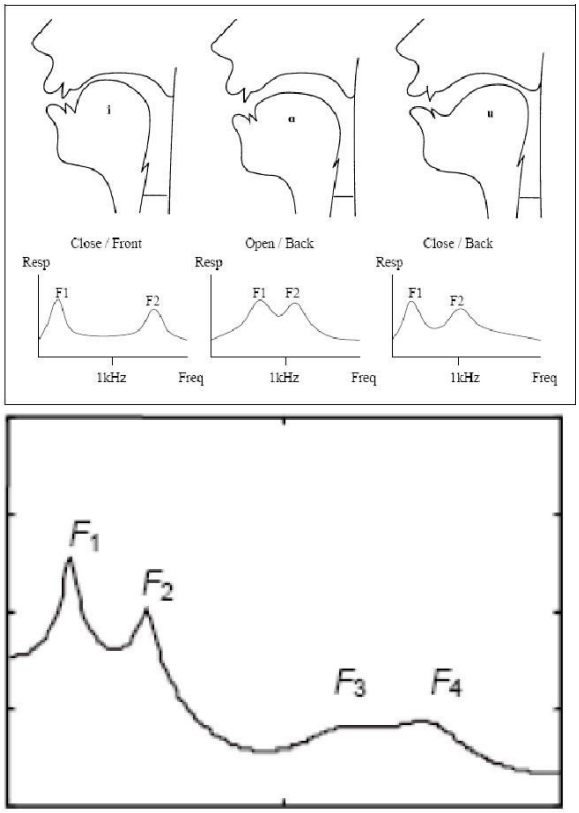


Fig. 2 Vocal tracts Frequency Response

Figure 2 exhibits the vocal tract frequency response. The x-axis represents the frequency scale band y axis represents the magnitude of the signal. Formants of the signal are classified as F1, F2, F3 and F4. Generally a voice signal will contain three to five formants. But in some voice signal up to four formants can be detected. LPC (Linear Predictive Coding) is used to obtain the formants of the voice signal. The LPC (Linear Predictive Coding) method is obtained from the word linear prediction. Linear predication term implies as the type of mathematical operation. This mathematical function which is used in discrete time signal estimates the future

$$\hat{x}(n) = -\sum_{l=1}^{P} a_l\, x(n-1) \quad (3)$$

The LPC will analyze the signal by predicting the formants present in the signal. Then, the formants effects are eliminated from the speech signal. The frequency and intensity of the remaining buzz is estimated. So by removing the formants the resonance effect will be eliminated. This process is called as the inverse filtering. The remaining signal is called as residue which does not contain any formants. To calculate the coefficients,

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

formants LPC is needed. The mean square error value is estimated by original value and predicated value. By reducing the error, the coefficients are detected with a

A. *Discrete Wavelet Transform*

The aim of a speaker recognition system is to anciently provide accurate and distinguishable individual properties of each speaker. Specie feature in individual speech are always the basis of a speaker recognition system. It is especially important, if a new speaker representation is introduced. It must be cleared out whether the representation carries suitable signal features or not. There are various methods of speaker recognition, where multivariate kernel density, Gaussian mixture models, arterial neural networks or support vector machine were used. There are numerous advantages in using voiceprints. Voice is a phenomenon that doesn't require the subject to be present. It can be recorded in some place and sent to another almost instantly. In addition, almost every user already has a microphone in PC or telephone whereas not manyca n aord a camera or anger scanner. Finger-prints, images or faces cans require to use more complicated devices. The goal of our research was to analyze, if the WFT can improve traditional speaker recognition methods. The system should be text independent and should be based on the speech characteristics such as accents, speaking styles and decencies. Wavelet packets were already tested for speaker variation. Perceptually motivated wavelet based methods were successfully tested for speech recognition .Another method of reduction of a wavelet decomposition tree for a speaker identication task was presented; however, described conclusions about their solutions being optimal are questionable.

B. Formant Estimation

Formant is one of the important components in the speech. The frequencies at which the resonant peak occurs are called as formant frequencies or simply formants. Vocal track frequency of the signal is analyzed to obtain the formant of the signal Discrete Wavelet Transform (DWT) is a revised version of the Continuous Wavelet Transform (CWT). CWT generates a huge amount of data which are compensated by the DWT. The basic operation of both DWT and CWT are similar however the scales used by the wavelet and their position are based upon the power of two. This is called as dyadic scales and there respective position are termed as dyadic stands for factor of two [9]. In real world application, most of the important feature of the signal lies in the lower frequency of the signal. To analyze the voice signal the lower frequency of the signal will provides the identity whereas the high frequency of the signal provides nuance to the signal. This is similar as the imparting flavor to the signal. In the voice signal if the higher mainly sense for non-periodic signal such as voice signal which has high frequency components dominates for short duration and low frequency components dominates for long duration. Small scales can be interpreted as "narrow" window. Using small scale fine details of the signal is analyzed. Vice versa, A large scales signal is used for "large window". Using large scale gross feature of the signal is analyzed. This property of the wavelet makes it more powerful and useful in detecting revealing hidden aspects of the data. Wavelet transform provides differ higher accuracy and the formants of the voice signal are obtained.

equation

$$f(n) = -a(1)x(n-1) - a(2)x(n-2) - a(3)x(n-3) \dots \dots (4)$$

SYSTEM IMPLEMENTATION

A. Variation System Implementation

In order to implement the system, certain method is employed to decompose the voice signal to its approximation and detail. Reorganization process is carried out by the approximation and the detail coefficients which are extracted from the signal. In the proposed method statistical calculation is carried out in the recognition phase which is mainly focused on the parameter. Four different type of statistical calculation is carried out with the coefficients. The statistical calculations which are carried out on the coefficient are mean, standard deviation, variance and mean of absolute deviation.

The wavelet that is employed with the system is the symlet 7 wavelet as that this wavelet has a very close correlation with voice signal. This is obtained through numerous trail and error. The coefficients that are extracted from the wavelet decomposition is the second level coefficient as it contain most of the correlated data of the voice signal. The data at the higher level contain very small amount of the data deeming which is not useful in the recognition phase. Hence for initial system implementation, the level two coefficients are used. The coefficients are further threshold to remove the lower correlation value. By using these coefficient value statistical computation is carried out which is used in the compression of voice signal with formant estimation and

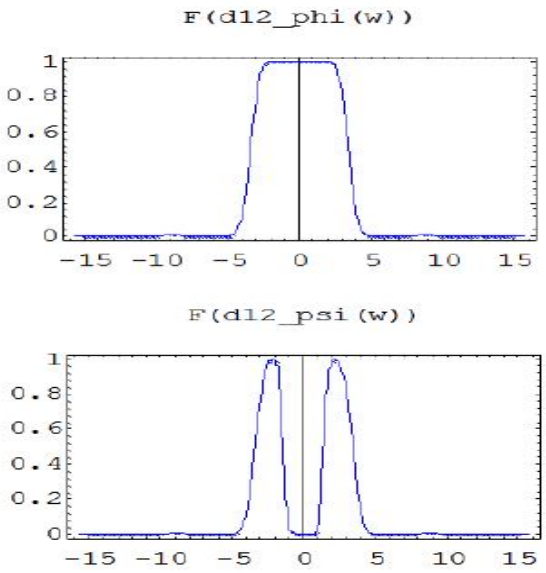
ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

wavelet energy. The entire extracted information act as the 'finger print' of the voice signal. The percentage of the voice signal is estimated by comparing the current values during the vocal folds vibration. This eect makes the glottal pulse spectrum to decrease with frequency in a faster way. However, as this glottal pulse will be later modied by the vocal tract, it is very hard to extract reliable speaker-independent information based on this issue.

Models of voice generation

Knowledge of how voice is produced and perceived by human being players an important role in speech technology systems. Speech is the result of activity in the various elements of the speaker's respiratory system. All of them contribute somehow to the final speech signal. Every block introduces some speaker-dependent information in the speech signal, however, only some of them can be exploited. Lungs The lungs are used for the vital function of inhalation and exhalation of air. In the speech production model they are the power source that supplies energy to the rest of the blocks in the systems. Inhalation is achieved by reducing the lung air pressure. This is possible thanks to the rib cage and the diaphragm. The rib cage is expanded during this process. The diaphragm, which is placed underneath the lungs, is lowered so the lungs are expanded. This pressure lowering causes air to rush in through the vocal tract and down the trachea into the lungs. Exhalation is opposite to inhalation. It is caused by an air pressure increase in the lungs. The volume of



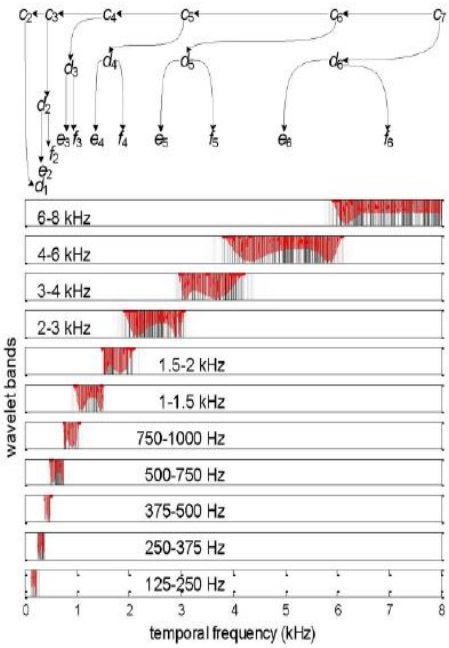
the chest cavity is reduced by contracting the muscles in the rib cage and lifting the diaphragm. This produces an airow from the lungs to the larynx through the trachea. Inhalation and exhalation always rhythmically follow the one to the other when breathing. However, during speaking short spurs of air are taken.

Larynx

The larynx, also called the \voice box" is a complex system of cartilages, muscles and ligaments. It has different functions such as closing the entrance to the lower respiratory system during swallowing. Since this kind of functions is not important for the speech production models, they will not be analysed. From the voice production point of view, the most important parts of the larynx are the vocal folds and the glottis. The vocal folds are two twin masses of freesh, ligament and muscle which stretch between the front and the back of the larynx.

HMM Wavelet-Fourier Transform

The standard transform used for speech signals analysis is the fast Fourier transform (FFT) which gives averaged representation of a signal in the frequency domain. Short Fourier transform is capable of carrying time frequency changes, however, analyzing windows creates artefacts.



ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

Figure 2: The frequency bands for speech signal analysis are perceptually motivated. They were chosen in the aim of representing the frequencies most important for a speech signal in several narrow sub bands and less important frequencies in wide sub bands. The general structure of decomposition tree was found in a process of optimization; however, the structure is its slightly smoothed version. The discrete wavelet transformation (DWT) belongs to the group of frequency transformations and is used to obtain a time-frequency spectrum of signal fs (n)g. This encourages us to use the DWT as an articial method of speech analysis. Dyadic frequency division makes the DWT much more compatible with the principles of the operation of human hearing system, equipped with subsystem for frequency analysis (to reveal the important information for the human speech recognition ability), than other methods. The wavelet transform (WT) is defined by formula.

$$\widetilde{s}_{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_{\infty}^{\infty} s(t)\psi\left(\frac{t-b}{a}\right) dt,$$

where a 2 <+ and b 2 <. The two arguments function es (a; b) represents wavelet spectrum of signal s(t). Parameter a, called scale, reversibly correlates with a frequency. Parameter b is a time translation. Function (t) is an arbitrary chosen wavelet and its example is presented in Fig.1.Formula (1) shows, that wavelet spectrum carries both, time and frequency representations. The events can be captured precisely, because the analyzing wavelet window into frequency. WT (1) has a simple physical interpretation: the analyzing function (t) is a exible time-scale window that automatically narrows at high frequencies and widens at low frequencies. A WT depicts information about the signal changeability in the time domain. This kind of analysis provides valuable information about voice irregularity in the time domain, according to frequency variations. It is an important property that the WT (1) has the form of a correlation operator. It enables us to apply the Fourier Transform (FT) to the WT and de_ne the new method of speech analysis. Let us consider the composition of two transforms. For a speech signal, the wavelet spectrum is calculated and next the FT is used to obtain

Their size varies from one person to another and in average it is around 15 [mm] long for men and 13 [mm] long for women. They can remain open to create unvoiced sounds or they can vibrate in order to produce voiced sounds during speech. During breathing, they remain open, allowing the air to flow into the lungs. Voiced sounds are characterized by the vibration of the vocal folds which for males due to anatomical characteristics such as the length and mass of vocal folds which is lower in the case of women. In the case of children pitch is even higher. Therefore, estimation of the pitch can be a good gender or age discriminator. In addition, pitch does not remain constant during speech. Some systems which use prosodic features take this pitch evolution into account despite the fact that it is relatively easy to imitate by an impostor. Some models have been created to model the airow velocity output at the glottis. The vocal folds are opened for a very short period. On the other hand, the breathy voice and the vocal folds remain open for a longer time. Both glottis responses show a pitch of 200 [Hz]. In the frequency domain, it can be seen that the longer the glottis remains open, the higher spectrum roll-o_ it shows. In addition, the spectrum contains a peak in every multiple of the fundamental frequency. Therefore, voice quality depends on the glottal pulse shape. Speaker individuality is also present in the quality of voice. This quality is lower in the case of a breathy voice, as the glottis is not almost close.

$$\widetilde{\widetilde{s}}_{\psi}(a,\omega) = \frac{1}{\sqrt{a}} \int_{\infty}^{\infty} e^{-j\omega b} \int_{\infty}^{\infty} s(t)\psi\left(\frac{t-b}{a}\right) dt db. \qquad (2)$$

FT is calculated with respect to the variable b, and the coecient a plays the role of constant parameter only. The wavelet-Fourier spectrum has two arguments. The one describes the frequency band, where its average value is proportional to 1=a and !, the second one, denotes the frequency in which the previous frequency appears in the signal.Formula (2) plays a role of WFT de_nitions and has small usefulness due to a large amount of calculations in numerical computing of integrals.To improve the computer calculations, DWT is used instead of (1) and FFT instead of FT in (2).For each wavelet (t) (see [11]) the scaling function '(t) is de_ned. These both functions have unique character, in a sense that each wavelet function (t) has only one scale function '(t).

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

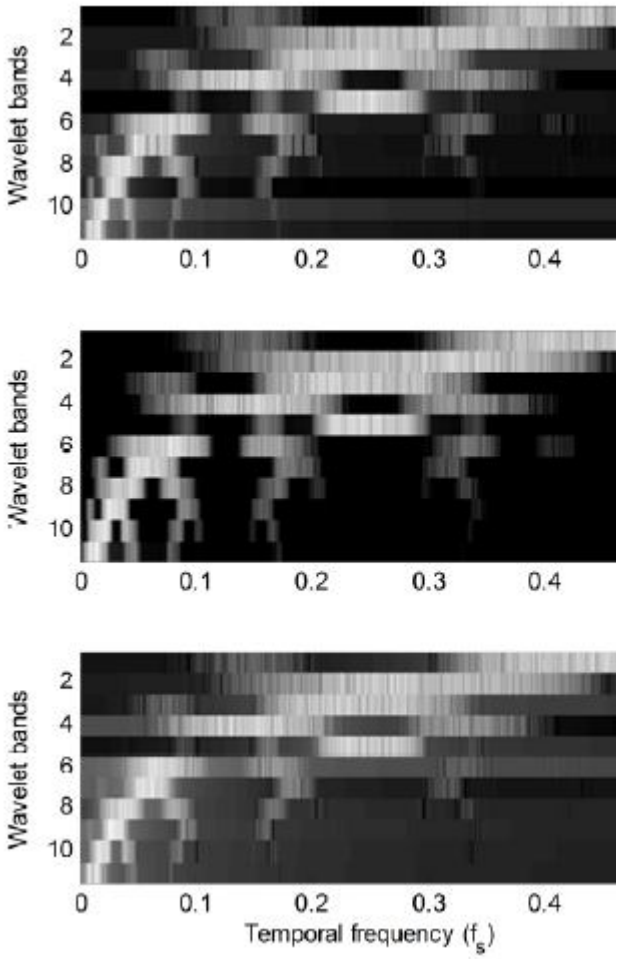


Figure 3: Example of DWFT spectra for three di_erent speakers and 11 resolution levels in a [dB] scale. DWFT allows easy and detailed analysis, similarly to traditional spectrograms but are more .

The wavelet that is employed with the system is the symlet 7 wavelet as that this wavelet has a very close correlation with voice signal. This is obtained through numerous trail and error. The coefficients that are extracted from the wavelet decomposition is the second level coefficient as it contain most of the correlated data of the voice signal. The data at the higher level contain very small amount of the data deeming which is not useful in the recognition phase. Hence for initial system implementation, the level two coefficients are used. The coefficients are further threshold to remove the lower correlation value. By using these coefficient value statistical computation is carried out which is used in the compression of voice signal with formant estimation and wavelet energy. The entire extracted information act as

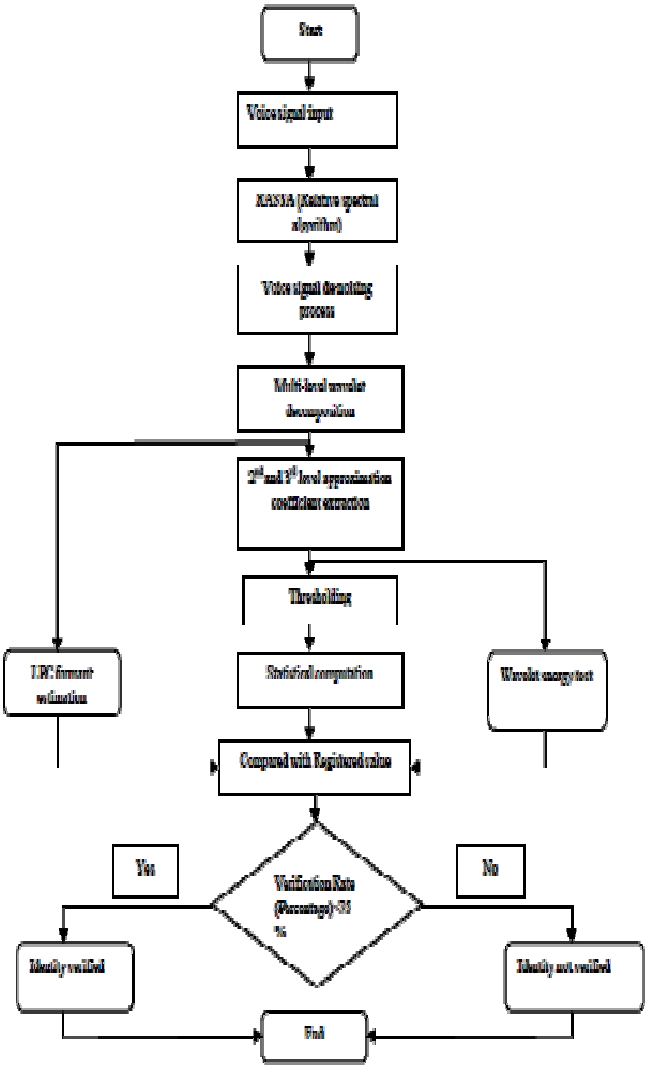


Fig. 3 Complete system Flow chart

Figure 3 shows the complete flow chart which includes all the components of the system which are important in signal values against the registered voice signal values. The percentage of the voice signal is given as: Verification% = (Test value/ Registered value) 100 Between the Resisted value and the tested value, whichever value is the highest one will be kept in the denominator and rest is in the numerator.

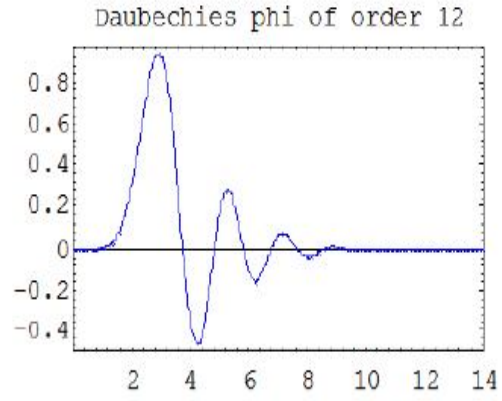
the 'finger print' of the voice signal. The percentage of the voice signal is estimated by comparing the current values

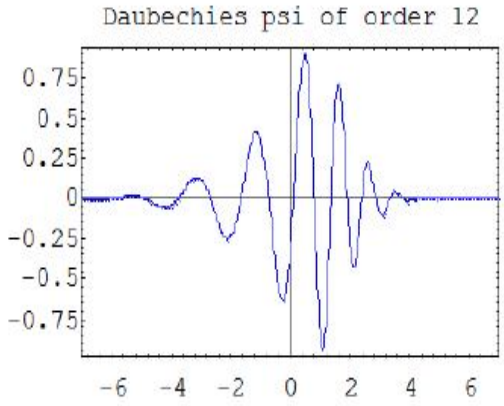
ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

B. Results

The complete systems which constitutes all the system components for the recognition methodology is one of the main reason for the high accuracy of the system. The percentage verification is set at an average value of 78.75%. The verification rate can be further increases or decrease by adjusting the percentage of the verification to higher to lower value. By substituting lower value the security of the system will decreases. it could jeopardize the accessibility rate of the system because for the certain level of tolerance is needed to the voice signal as it changes with internal factor and external factors.





C. GUI (Graphical User Interface)

Figure 4 shows the GUI(Graphical User Interface) implementation for the Voice Registration Section GUI enables the user to register an individual's voice signal using the pre-loaded voice signal that are saved in the program. The STEP 1 panel shows the pre loaded voice signals contained in the program. The voice signals are taken from the noise free and clean environment. The user can select the available voice signal by using pop-up menu. Show Voice Plot button enables the user to view the voice plot in the graph as shown in panel. The STEP 2 panel contains the function that enables the user to de-noise the signal and view the de-noised signal in graphical representation as shown in plot. The extract coefficient button enables the user to view the DWT (Discrete Wavelet Transform) coefficient detail and approximation plot. The STEP 3 panel shows the extracted coefficient plot which is the 2nd level approximation and detail and the 3rd level detail and approximation. The STEP 4 panel shows the recognition method of the program. The compute value button performs the statistical computation on the 2nd level approximation and 3rd level approximation and displays these values. The wavelet energy and formant value of the voice signal is calculated and shown. The STEP 1 shows the pre-load voice signal contained in the program. These voice signals are recorded from different individual saying their own name. Using pop-up menu user can select the particular voice. The plot voice signal button enables the user to see the graphical representation of the voice. The STEP 2 panel shows the recognition method of the program. The compute value button executes the statistical computation of the 2nd level approximation and the 3rd level approximation and displays these values. The formant values and the wavelet energy of the voice signal is estimated and shown.

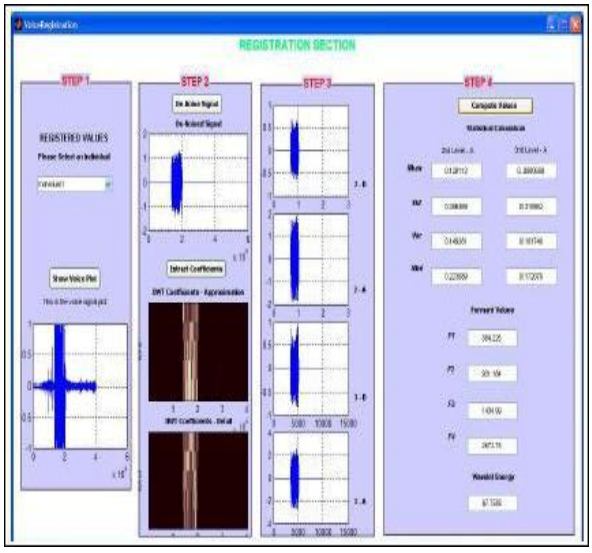


Figure 4: Voice Registration GUI

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 319 - 327 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp61.pdf

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

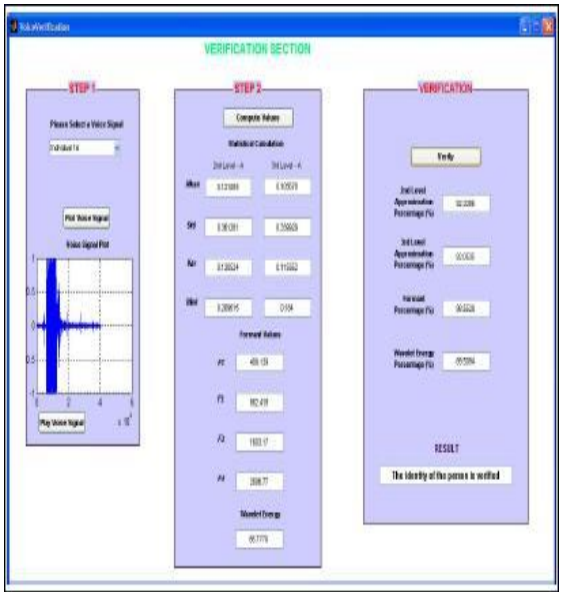


Figure 5: Voice Verification GUI

The VERIFICATION panel shows the verification process of the system. The overall percentage value of the statistical computation, formant values and the wavelet energy are displayed

CONCLUSION

The voice recognition using Wavelet Feature Extraction employ wavelets in voice recognition for studying the dynamic properties and characteristics of the voice signal. This carried out by calculating 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 dependent 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. GUI is build for the user. In order to make it easy for the user to understand the operation step by step take place in the wavelet transform. By using the fifty preloaded voice signal from five individuals verification is carried out and the accuracy of the system is approximately 80% is achieved. The system can be enhanced further by using the advance pattern recognition technique such as Neural Network or Hidden Markov Model (HMM) for efficiency.

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.