International Journal of Engineering & Technology, 7 (4.6) (2018) 285-290

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Novel User Interface for Text Dependent Human Voice Recognition System

Ramadevi P.

Associate Professor, Department of ECE, Vardaman College of Engineering, Hyderabad, Telangana, India
**Corresponding author E-mail:p.ramadevi@vardhaman.org*

Abstract

In an effort to provide a more efficient representation of the speech signal, the application of the wavelet analysis is considered. This research presents an effective and robust method for extracting features for speech processing. Here, we proposed a novel user interface for Text Dependent Human Voice Recognition (TD-HVR) system. The proposed HVR model utilizes decimated bi-orthogonal wavelet transform (DBT) approach to extract the low level features from the given input voice signal, then the noise elimination will be done by band pass filtering followed by normalization for better quality of a voice signal and finally the formants of a train and test voices will be calculated by using the Additive Prognostication (AP) algorithm. Simulation results have been compared with the existing HVR schemes, and shown that the proposed user interface system has performed superior to the conventional HVR systems with an accuracy rate of approximately 99 %.

Keywords: *Additive Prognostication (AP); band-pass filtering; feature extraction; human voice; recognition rate; Wavelet decomposition/reconstruction tree.*

1. Introduction

In our regular day to day existences the audio flag particularly the voice signal has gotten to be one of the real part, since it can be utilized as a one of the real instrument for communicating each other. Be that as it may, by utilizing changed handled because of innovative progression, used in different , for example, numerous applications these discourse preparing assumes an imperative part, for example, discourse acknowledgment, voice communication. Discourse acknowledgment is the procedure of consequently extricating and deciding etymological information passed on by a discourse signal utilizing PCs or electronic circuits. Programmed discourse acknowledgment techniques, examined for a long time have been mainly gone for acknowledging translation and human PC association systems. The main specialized paper to show up on discourse acknowledgment has from that point forward heightened the scrutinizes discourse as of late developed, despite the fact that they stay just of constrained use.

2. Voice Recognition

Most discourse acknowledgment systems can be characterized by following classifications:

2. 1 Speaker Dependent versus Speaker Independent

Discourse acknowledgment framework prepared perceive discourse standout worked only solitary individual, henceforth financially reasonable. On the other hand, framework autonomy difficult accomplish, discourse acknowledgment have a tendency end up prepared, bringing about subordinate.

2.2 Isolated versus Constant

In detached discourse, delays each consistent discourse talks persistent perhaps almost the middle. Secluded discourse acknowledgment systems are anything but difficult to work, as it is minor to figure out where single word closures and another begins, and every word has a tendency to be all the more neatly and unmistakably talked. Words talked in ceaseless discourse then again explanation impact,
adjusted preparing discourse framework troublesome, might numerous conflicting .

3. Related work

There are many researches those have been done with human voice recognition from the past decades. Most of them had done with speaker or speech recognition systems with various algorithms like LPC, MFCC and HMM. LPC will approximate the envelope of a voice signal range. It is an excellent tool for audio signal processing and speech processing.

However, the recognition rate will be very poor with these algorithms for various types of voice signals such as male, female, children and old age persons.

3.1 Linear Predictive Coding (LPC)

The LPC technique is gotten from the word straight expectation. Direct forecast as the term infers is a sort of numerical operation. This scientific function which is utilized as a part of DTS gauges the future qualities depending on a direct role of past specimens [8].

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

$\hat{x}(n)$is the expected value and $x(n-l)$ is the previous value. By expanding this equation

$$\hat{x}(n) = -[a_1 x(n-l) - a_2 x(n-l) - a_3 x(n-l) \ldots.$$

The LPC can investigate the sign on an assessment or foreseeing the spectral envelopes. At that point, discourse signal expelled the spectral envelopes impacts. For rest of the buzz power and frequency are evaluated. Expulsion of envelopes from the voices sign kills reverberation impact. The procedure is known as backwards filtering. Staying signal without the envelope is known as build-up. With a specific end goal to gauge the spectral envelopes, "coefficients of the LPC" are required. Evaluation of these coefficients will be done by mean square error between anticipated sign and first flag. After error reduction, coefficients will be recognized with superior exactness and voice sign envelopes will be acquired.

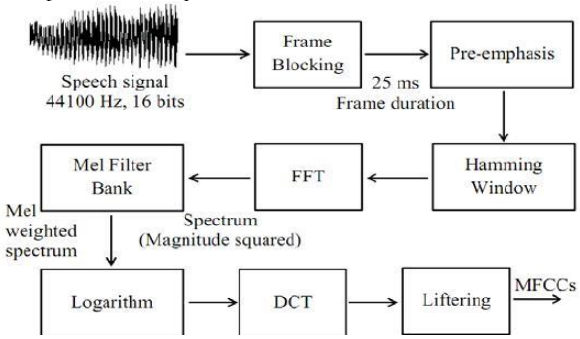


Fig.1: Block diagram of LPC based HVR model

3.2 Mel-frequency Cepstral Coefficients

These are the coefficient that have been made up with Mel-frequency cepstrum (MFC) and is a typical derivative of cepstral representation of a voice signal. The cepstrum is also known as a method that will be utilized to process an audio or voice signal. The major difference between these approaches is that in the MFC, Mel-scale will be used for the equal spacing of the frequency bands, due to this nature it will give more exact response of human auditory system than the frequency bands that have been spaced linearly, which occurs during the regular cepstrum. This warp nature of frequency will allow for improved depiction of audio signal as shown in figure 1.
The procedure of deriving MFCCs from a given voice clip is as follows:

- First, calculate the Fourier transform of (a windowed excerpt of) input signal to get the spectrum powers.
- Now, utilize the Mel scale to map the obtained spectrum powers using triangular overlapping windows.
- Now, at each Mel frequency calculate the powers logarithms

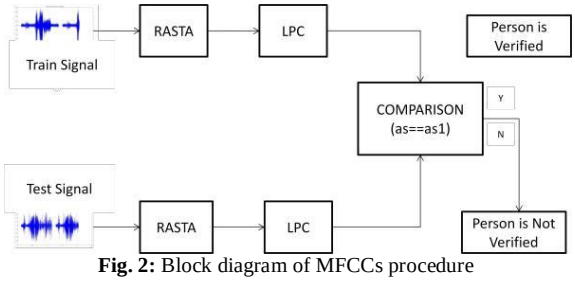


Fig. 2: Block diagram of MFCCs procedure

- Finally, compute the discrete cosine transform (DCT) for the above obtained log powers. The resulting spectrum amplitudes are the MFCCs

4. Proposed TD-HVR model

The proposed technique for the acknowledgment stage is the measurable figuring. Four unique sorts of factual estimations are done upon these coefficients. Measurable estimations being done difference . Utilized for the framework bi-orthogonal wavelet (BW) which has nearby connection with voice signal which is resolved by various assessments. Coefficients separated by wavelet disintegration procedure are the second level coefficients. These will hold the majority of voice related information. Information on more elevated amounts holds next to no information considering it not viable for the acknowledgment stage. Thus to begin framework execution, the second level coefficients will be utilized. These are more edge for evacuating short relationship amounts. By utilizing those measurable calculations are completed. Factual calculation is utilized as a part of voice flag examination with the formant estimation and wavelet vitality. Removed information performs as "unique mark" in favor of voice signal. Check rate will be figured by contrasting present valuable signal qualities aligned with enlisted those of voice signal.

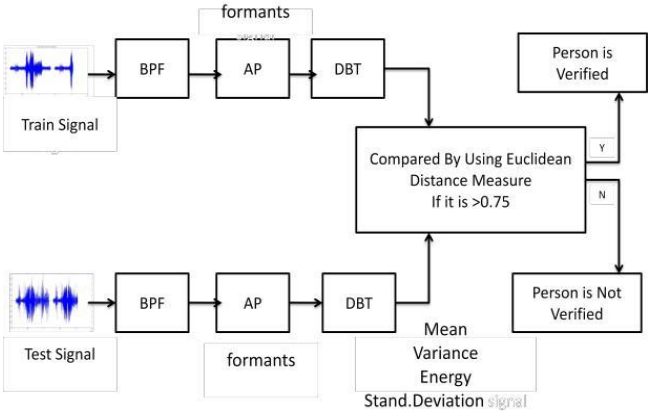


Fig.3: Block diagram of proposed TD-HVR model

Algorithm:

Step1: First, we will apply the band pass filter followed by normalization to the trained voice signal to obtain the de-noised voice signal

Step2: The signal obtained in step1 will be given as an input to the AP algorithm, which will be utilized to estimate the formants of trained voice signal

Step3: Now, apply decimated bi-orthogonal transform to extract the low level features and calculate the statistical parameters such as mean, variance and standard deviation to the obtained low level coefficients

Step4: Now, calculate the wavelet energy for the low level and high level coefficients of trained voice signal

Step5: Apply the step1 to step4 for the testing voice signal and form a train and test feature vectors that consists of statistical parameters, low level features and estimated formants.

Step6: Now, the recognition will be calculated by using the verification % given below:

Verification % = (Test value / Registered value) x100

Step7: Finally, if the verification percentage is above 75% then the person identification will be successful otherwise it will displays that the person is not found

4.1 Wavelet Analysis

Fundamental thought proposition utilize separating viewed as moderately sign handling contrasted with different strategies or techniques as of now utilized STFT [1] and [2] present techniques utilized as a part of the field of sign preparing. However because of serious constraints forced investigating considers inadequate dissecting unpredictable 'Flags'. For example, voice sign [3] and [4]. Fourier Transform (FT) contains some disadvantage as this works out for "stationary signals" alone, with time period independent. As FT is valid for the complete signal and not for signal sections. Non-stationary signal is not transformed by FT. Another disadvantage is that with the FT occurrence of a particular event cannot be predicted. To overcome the disadvantage in FT, during 1946 a novel method known as “Windowing” was coined by Dennis Gabor. Windowing could be useful to investigate even a minor signal section. This version is known as "Short-Time Fourier Transform (STFT)". In STFT, sample is resolved as time and frequency. In STFT, the window is fixed and is independent from time period of the signal. The frequency content cannot be predicted in time intervals. To defeat these drawbacks of STFT, a *wavelet* technique is established. In this window size is variable. Wavelet analysis permits long time intervals usage where more precise low-frequency information is required and shorter regions where high-frequency information is needed.

Wavelet method is utilized for separating voice signal specifications by handling information at various levels. Wavelet method controls these levels to provide superior relationship during recognizing different 'frequency components' in this sign. The specifications will be additionally prepared keeping in mind this end goal is to build the voice acknowledgment framework. Separating voice signal specs has no constrain over the abilities of this method for any specific function only, yet this unlocks this way for an extensive variety of potentials for various purposes could profit by voice extricated methods. Functions, for example, discourse acknowledgment framework, discourse to content interpreters, and voice based security framework are a portion without bounds systems that can be created.

4.2 Decimated Bi-orthogonal Transform (DBT)

Decimated bi-orthogonal transform is very much utilized for multi resolution analysis because of its multi scaling functionality i.e., two scaling functions to generate wavelet channel banks for disintegration and remaking separately. It will give more viable disintegration coefficients because of its multi scaling property. In the case of orthogonal, we have one hierarchy of approximation spaces $_{-1} \subset \subset _{+1}$ and an orthogonal decomposition

$$V_{j+1} = V_j \oplus W_j \tag{1}$$

which leads us to use two filter sequences and for decomposition and reconstruction. Hence, we need to construct two different wavelet functions and two different scaling functions.
Let $f_k , g_k \in H . if \langle f_j, g_k \rangle = \delta_{jk}$ Then we will say that the two sequences are biorthogonal.
Now, our aim is to build two sets of wavelets

$$\psi_{j,k} = 2^{\frac{j}{2}} \psi\left(2^j x - k\right) \tag{2}$$

$$\tilde{\psi}_{j,k} = 2^{\frac{j}{2}} \psi 2^j x - k \tag{3}$$

To do so, we need four filters$g, h, \tilde{g}, \tilde{h}$ i.e., two sequences to be act as decomposition sequences and two sequences as reconstruction sequences. For example, if $c_n^1$is a data set, it will be decomposed as follows:

$$c_n^0 = \sum_k h_{2n-k} c_k^1 \tag{4}$$

$$d_n^0 = \sum_k g_{2n-k} c_k^1 \tag{5}$$

And the reconstruction is given by

$$c_l^1 = \sum_n \tilde{h}_{2n-l} c_n^0 + \tilde{g}_{2n-l} d_n^0 \tag{6}$$

We can achieve perfect reconstruction by following some conditions given below:

$$g_n = (-1)^{n+1} \tilde{h}_{-n}, \ \tilde{g}_n = (-1)^{n+1} h_n$$

$$\sum_n h_m \tilde{h}_{n+2k} = \delta_{k0}$$

Now consider that $\phi(x)$and $\tilde{\phi}(x)$ are two scaling function with their own hierarchy of approximation spaces, then we will generate function of wavelet in a method of analogous to the orthogonal case. We now define the scaling function as follows:

$$\phi(x) = \sum_n \sqrt{2} \sum_n h_n \phi(2x - n) \tag{7}$$

$$\tilde{\phi}(x) = \sqrt{2} \sum_n \tilde{h}_n \phi(2x - n) \tag{8}$$

So, finally the bi-orthogonal wavelet functions can be defined as follows:

$$\psi(x) = \sqrt{2} \sum_n g_n \phi(2x - n) \tag{9}$$

$$\tilde{\psi}(x) = \sqrt{2} \sum_n \tilde{g}_{n+1} \tilde{\phi}(2x - n) \tag{10}$$

5. Simulation results

The experiments have been done with graphical user interface in MATLAB environment. We considered various tested and trained voice signals in real time environment i.e., recorded voice has been taken directly and converted into the format in such as way it will be read by MATLAB for the better analysis of HVR system. GUI model of proposed voice registration system has been shown in fig 4, 5, 6, 7 and 8. Verification has been shown in fig 9, 10, 11, 12 and 13. Fig 14 and automatically it has shown the message that the person is not verified; Finally, LPC achieved 66.66% accuracy, MFCC achieved 75% and our proposed HVR model achieved almost 90.9% accuracy.

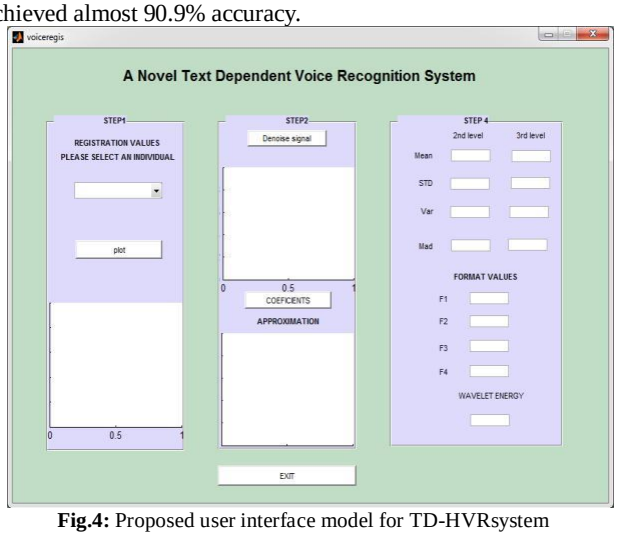


Fig.4: Proposed user interface model for TD-HVRsystem

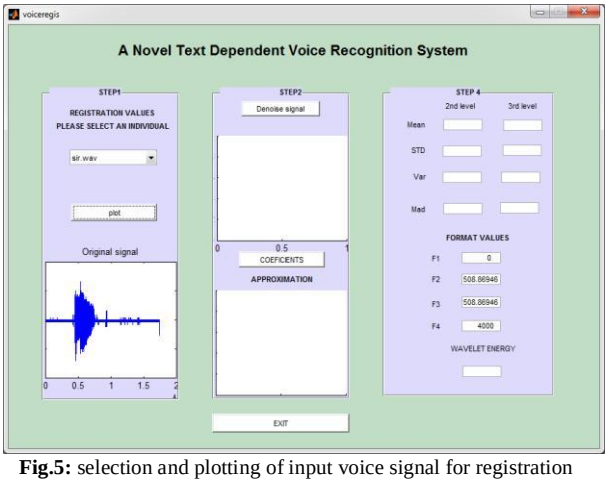


Fig.5: selection and plotting of input voice signal for registration

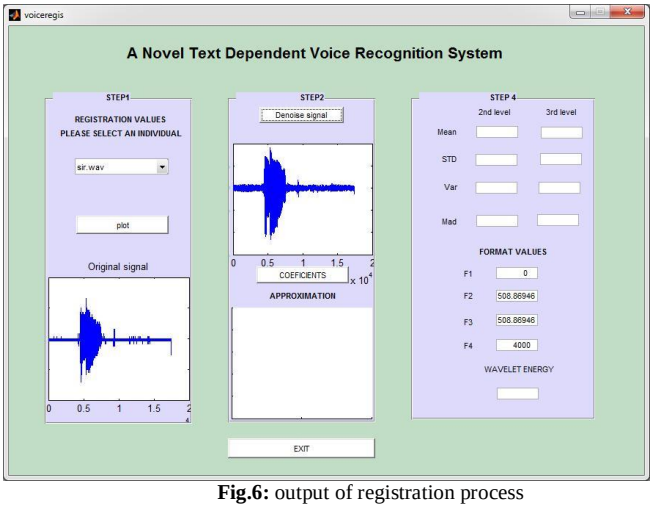


Fig.6: output of registration process

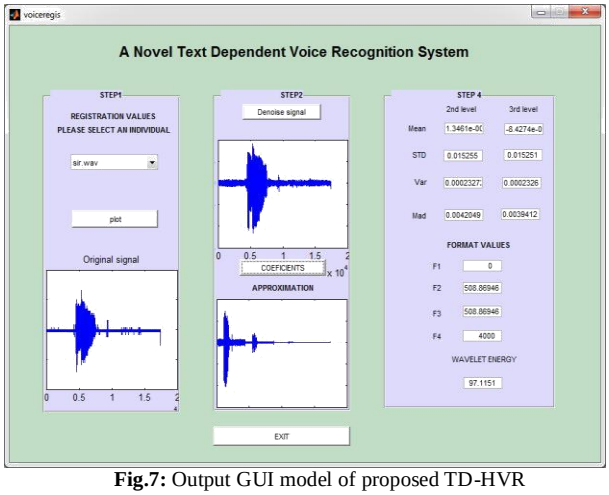


Fig.7: Output GUI model of proposed TD-HVR

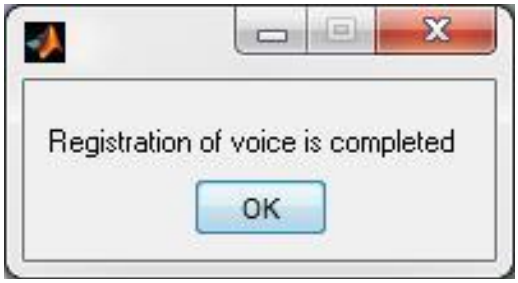


Fig.8: Registration process completed

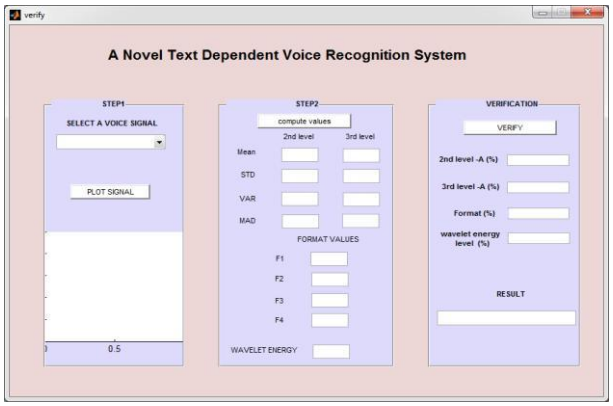


Fig.9: Proposed GUI model for Voice verificationprocess

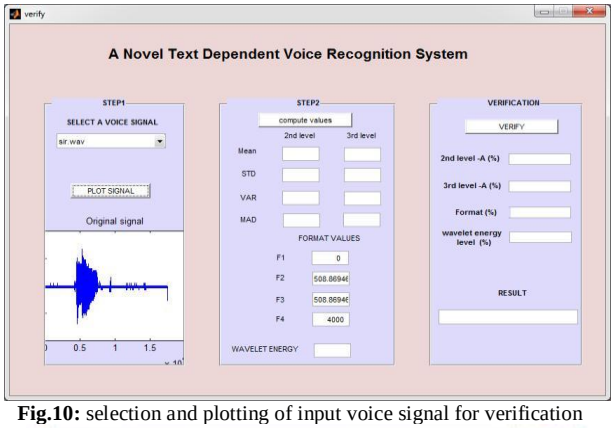


Fig.10: selection and plotting of input voice signal for verification

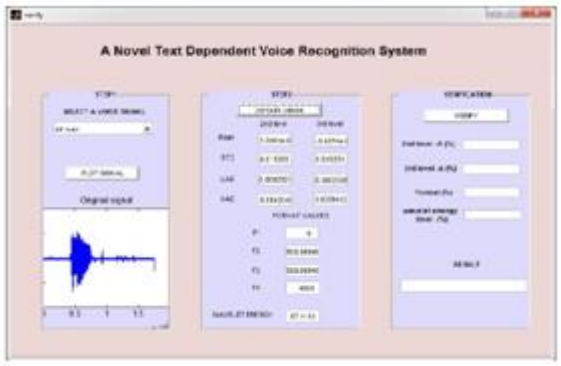


Fig.11: GUI model of verification after DBT

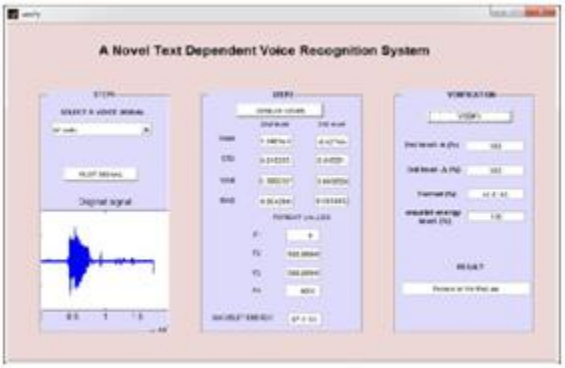
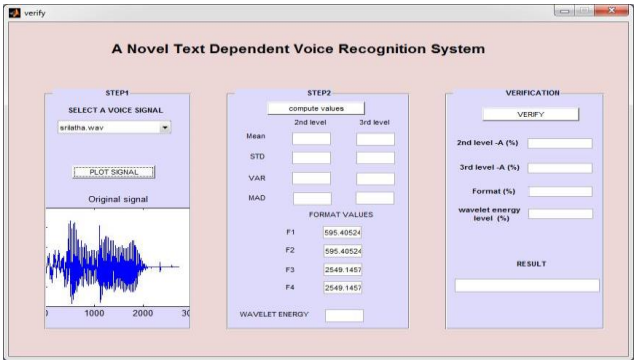
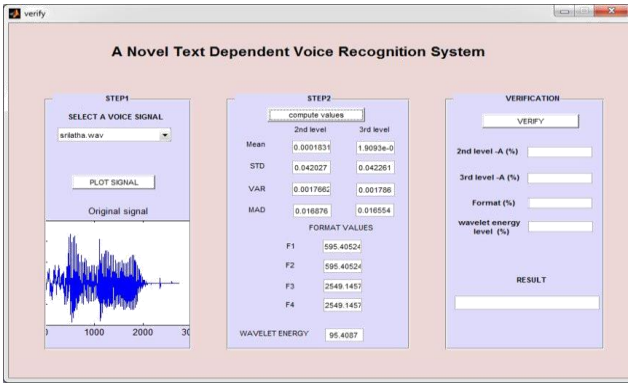


Fig.12: Person verification has been completed

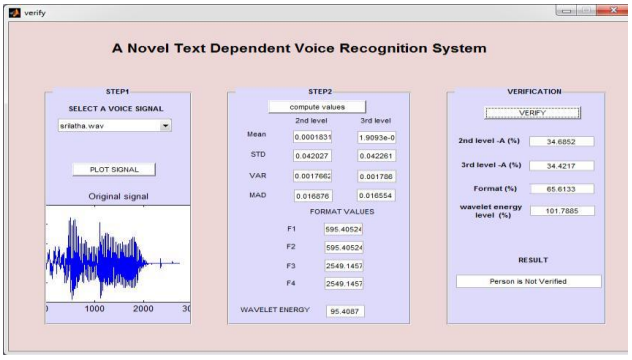
Fig. 13: message box after completion of verification process



(a)



(b)



(c)

(d)

Fig.14: (a), (b), (c) and (d) GUI models for unregistered voice signal

6. Conclusions & future work

We proposed a novel user interface model for text dependent human voice recognition (TD-HVR) system with various voice signals in real time environment using MATLAB tool. Proposed

system also employed to examine the uniqueness for any entity support for personal speech signal with "statistical parameters computation", estimating the formants and wavelet energy for low level features extracted using decimated bi-orthogonal transform. After a thorough analysis a precision rate of nearly 90 % is reached conventional schemes 66.66% and 75% respectively greater extent and it can be used as a HVR tool in real time applications. Furthermore, the same can be implemented by interfacing the MATLAB with some hardware support packages such as arduino, Raspberry Pi for voice operated robotic applications

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, ManikRaina, Subhas K.Ghosh, Signal Separation using Time Frequency Representation, Honeywell Technology Solutions Laboratory.
[5] Amara Graps, An Introduction to Wavelets, Istituto di Fisica dello Spazio Interplanetario, CNR-ARTOV BraniVidakovic and Peter Mueller, Wavelets For Kids– A Tutorial Introduction, Duke University.
[6] O. Farooq and S. Datta, A Novel Wavelet Based Pre Processing For Robust Features In ASR.
[7] GiulianoAntoniol, 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.
[8] Michael Unser, Thierry Blu, IEEE Transactions on Signal Processing, Wavelet Theory Demystified, Vol.51,No. 2,Feb'13.
[9] C. Valens, IEEE, A Really Friendly Guide to Wavelets,Vol.86, No. 11, Nov 2012.
[10] James M. Lewis, C. S Burrus, Approximate CWT with An Application To Noise Reduction, Rice University, Houston.
[11] Ted Painter, Andreas Spanias, IEEE, Perceptual Coding of Digital Audio, ASU.
[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] Rami Al-Hmouz, Khaled and Ali, "Multimodal Biometrics Using Multiple Feature Representations toSpeaker Identification System", InternationalConference on Information and Communication Technology Research (ICICTR), 2015.