International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST229317

Speaker Verification using CWT DWT Transform and Neural Network

Dr. Vinodpuri Rampuri Gosavi, Dr. Manju D. Pawar, Dr. Ganesh S. Sable
G. S. Mandal's, Maharashtra Institute of Technology (An Autonomous Institute) Aurangabad, Maharashtra India

Article Info
Volume 9, Issue 3
Page Number : 125-130

Publication Issue
May-June-2022

Article History
Accepted : 01 May 2022
Published : 15 May 2022

ABSTRACT
In this Present study, the technique of wavelet transform and neural network were developed for speech based text-dependent and text0independent speaker identification. 390 feature were fed to feed-forward back propagation neural network for classification The function of feature extraction and classification are performed using wavelet and neural network system. The declared result shows that the proposed method can make an effective analysis with average identification rate reaching 98%. The best recognition rate selection obtained was for FFBPNN (Feed Forward Back Propagation Neural Network).
Keywords : Wavelet Transform, Neural Network, FFBPNN, Feature Extraction, Database.

I. INTRODUCTION

The Speaker Identification is the process of utterance speaker verification, on the other hand, is the technology of accepting or rejecting the identity claim of a speaker . Over the last four decays many solutions of speaker recognition have been appeared in literature.

The algorithm for pattern classification, was motivated by Patterson‟s and Womack‟s and wee‟s proof‟s that the mean Square Error (MSE) Solution of the Pattern classification solution gives a minimum mean square error approximation to Baye‟s discrimination weighted by the probability density function of the sample. All audio techniques start by converting the raw speech signal into sequences of acoustic feature vector carrying distinct information about the signal. This feature extraction is also called „front-end" in the literature. The most commonly used acoustic vectors are Mel Frequency Coefficients (MFCC), Linear Prediction Cepstral Coefficient (LPCC) Coefficient. All these features are based on the Spectral information derived from a short time windowed segment of speech signal. One of the most Common short-term spectral measurements currently used are Linear Predictive coding (LPC) derived Coefficients and their regression Coefficients. A Spectral envelope reconstructed from a truncated set of cepstral coefficients is much smoother than one reconstructed from LPC coefficients. Therefore it provide a stabler representation from one repletion to another speakers utterances" As for the regression coefficients, typically the first and second order coefficients are extracted at every frame period to represent the spectral dynamics.

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

These coefficients are derivatives of the cepstral coefficients and are respectively called the delta and delta-delta cepstral coefficients. Text dependent methods are usually based on template matching techniques. In this approach, the input utterances is represented by a sequences of feature vectors, generally short-term spectral feature vectors. The time axis of the input utterance and each reference template or reference model of the registered speakers are aligned using a dynamic time warping (DTW) Algorithm and the degree of similarity between them, accumulated from the beginning to the end of the utterance, is calculated. The Hidden Markov Model (HMM) can efficiently model statistical variation in spectral features. Therefore HMM based methods were introduced as extensions of the DTW-based methods, and have achieved significantly better recognition accuracies, One of the most successful text-independent recognition methods based on vector quantization (VQ), In this methods, VQ Codebooks consists of a small number of representative feature vectors are used as an efficient means of characterizing.

Speaker-specific features. A Speaker-specific codebook is generated by clustering the training feature vectors to each speaker. In the recognition stage, an input utterance is vector-quantized using the codebook of each reference speaker and the VQ distortion accumulated over the entire input utterance is used to make the recognition decision. Temporal variation in speech signal parameter over the long term can be represented by stochastic Markova transition between states. Therefore, methods using an ergodic HMM, Where all possible transition between states are allowed, have been proposed. Speech segments are classified into one of the board phonetic categories corresponding to the HMM states. After the classification, appropriate features are selected.

It has been shown that a common ergodic HMM method is far superior to a discrete ergodic HMM method and that a continuous ergodic HMM method is as robust as

VQ based method when enough training data is available.

A method using statistical dynamic features has recently been proposed. In this method, a multivariate auto- regression (MAR) Model is applied to the time series of cepstral vectors and used to characterize speakers [26]. In this paper, the wavelet Transform based speaker recognition system is proposed. This system is divided into two main blocks, signal enhancement by feature extracting and identification. In the first block use Adaline as neural network to enhance each sub-signal that produced by the DWT. This system depends on DWT generates the desired sub-signals to the neural net, This means multiple input will be applied to the neural net depends on selected level. The same process is applied for noisy signal, The output of the neural net will be back to the Inverse DWT to be reconstructed enhanced signal. The aim of this method is to filter the speech signal from the noise in selected sud-signals of distinct frequency sub-band. This assist greatly in eliminating special frequencies and can preserve other frequencies that are essential for speech recognition. IAn. Discrete Wavelet Transform and Adaline Enhancement speaker identification two blocks are applied: the first is WGD block, where three continuous wavelet transform (CWT) sub-signals are used of different pass band of frequency, high average and low . Then stastical functions are used to extract gender features. Sharp threshold between male and female is achieved. The second step is speaker identification using the enhanced sub-signal of suitable level that must be on speakers own features frequency, depending on his anatomical structure of his own vocal tract and other working parts through speaking process is used for feature extraction. And finally neural network is used for classify the features.

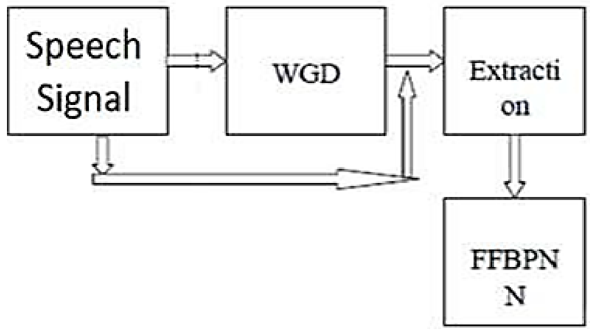

Fig.1

Block dig. of Proposed Method

II. PROPOSED METHOD

In My paper wavelet transform based identification system is presented. This particular system based on two main blocks shown in fig.1. In this first block speech signal is enhanced by Continuous Wavelet Transform method. The Second block contains the feature extraction method by Discrete wavelet transform and classification using FFBPNN(Feed Forward Back Propagation Neural network). Feature extraction method in the second block is divided into two steps: Wavelet Gender Discrimination and Feature Extraction, which assists in discrimination the speech signal into two classes (male and female) that makes the feature extraction by wavelet and is more efficient.

III. Method

Wavelet is the mathematical function that decomposes data into different frequency components and then study each component with a resolution matched to its scale. Wavelets were developed independently in the field of mathematics,quantum physics, electrical engineering interchanging between these fields during the last twelve years have led to many new Wavelet applications such as turbulences, image compression, human vision, radar.

In this, we use Adaline as neural net to filter each sub- signal or one that produced by DWT.

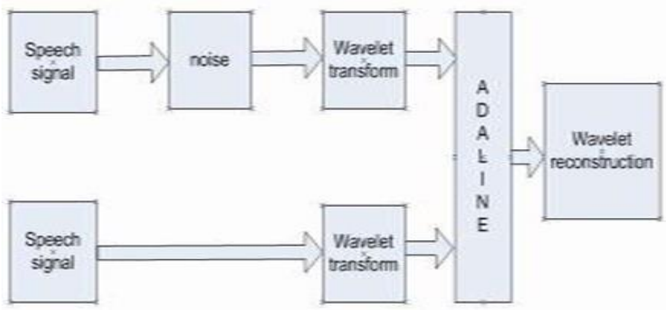

Fig.2 Practical noise reduction using neural Net and wavelet transform.

The ain of this method is to filter the speech signal from the noise in selected sub-signals of distinct frequency sub band, which can preserve other frequencies that is essential for speaker recognition system.

B. Gender Discrimination By CWT

The main advantage of using WT for signal analysis is its ability to decompose any given signal into a multiscale presentation. This enables the analysis of a given signal on different frequency bands, and helps in defining the most essentials scales of that signal. Figure CWT or DWT can be used in such analysis we choose the CWT, Since it provides more data per scale if compared to DWT . The extra –redundant data resulting from CWT Analysis will be useful in the computation of the standard deviation for the selected scales. The most essentials scales are 1,5,10, The mean of these scales $\mu\sigma$ is denoted as,

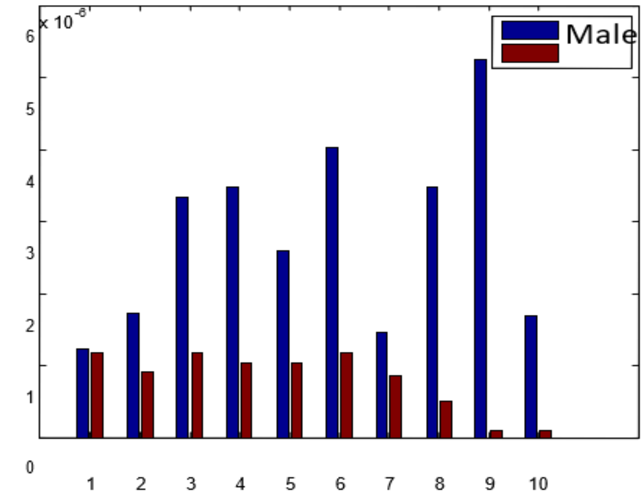

Fig.3 Gender Discrimination using μσ

From the above result, the following remarks can be concluded:

μσ = σ1+σ5+σ10 / 3

It is noticed that μσ of male speech signal is significantly larger than that of the male. As a matter of fact, this can be caused by the male vocal tract influence signals were used. Male and female input signals wer3e) The above result suggests that CWT is more suitable extracted from the speech data-base recorded by using windows recorder and use to compute the classifier tool to be used for non stationary signal analysis as compared to fast Fourier Transform.

The equation no. (1) is used for gender discrimination, on the CWT function.

The parameter μσ has lead to more accurate result2s) The CWT Scales of female signal are significantly in the male and female identification. The CWT based methods having varieties of input different because CWT can accurately locate sub-signal of different frequency.

Threshold for male and female cases, In this pape4r) It can be shown that μσ > 1.4 for male and μσ < 1.4 for female. This gives a good classification measure accounting for speaker gender identification.

We have used different 10 speech signals for 4 male and 4 female, so the total 80 signals are used, In figure 4, computed μσ for male and female signals.

C. Wavelet Feature Extraction

After gender discrimination, speech signal is given to the stage of feature tracking method by DWT feature extraction, In this stage the speech signal is decomposed into discrete wavelet transform sub signal (d1,d2,d3......dj,sj),where each of these sub-signals is generated by mallet"s algorithm of particular level 1,2,...J. This accomplished by convolution the signal with mother wavelet function to achieve high-pass sub- signal of speech signal, and SJ, which is accomplished by convolution with father wavelet function to achieve low pass sub-signal of the speech signal. The speaker vocal tract frequency which is known as low frequency is contained in d5 or s5, but s5 has better capability of signal energy conserving, PSD is used to concentrate the signal energy and the more clear features can be demonstrated .where each person has own number of different feature bandwidth. In this case, the system can cancel the trail if no match is achieved by neural network.

IV. RESULT AND DISCUSSION

Tested speech signals Database were recorded via PC-Sound card, with spectral frequency 4000 HZ and sampling frequency 8000 HZ, over about 2 sec. Time duration.

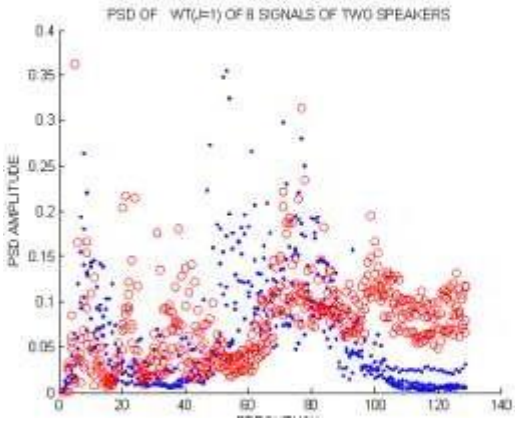

Fig. 4 The Effect of WT and PSD, of 10 signals (for two persons)

Each speaker recorded „Jordon Kingdom" that was recorded one time by one speaker. Total 39 individual speakers, including 20 individual Females and 19 individual females, spoke these „Jordon Kingdom" words and texts for training and testing phases. The total number of tokens considered for training and testing was 390. From above speech signal discretion, we can notice that presented recognition system is „text-dependent" system.

Fig. presents the effect of wavelet transform at classification due to separating the users features. These features assist greatly in discrimination the tested signals from the models speech signals stored in the system.

For speech signal identification via verification Neural Network are studied: Feed Forward Back Prorogation Neural Network. Table 1 shoes the results of FFBPNN by newff Matlab function. We can notice the ability of identification based on this algorithm.

After that the signal was classified by FFBPNN to achieve 98%. In other cases Wavelet Transform is more robust to noise separating of very noised signals.

For same purpose we create a FFBPNN NNT of input matrix of (P=[t]), t1 is PSD of WT of user one and binary target (T=[001] For Male, and Target T=[110] for Female. We train the network to classify t1 as 0 and same t1 as 1 and simulate other signal t2 of different speaker, and plot input/target vectors for Female in fig.no.

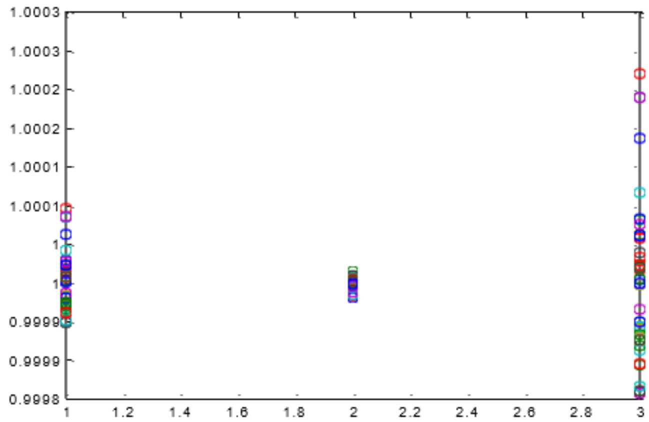

Fig. 5 : Test-Target output (Female) plot

Fig. 6 illustrate the plot of Regression is the correlation between output and Target. It is measure how well the variation in the output is explained by the Target. If this regression (R) value is equals to 1, which indicated a good fit. In my experiment this value is very much close to 1 which indicates a good fit.

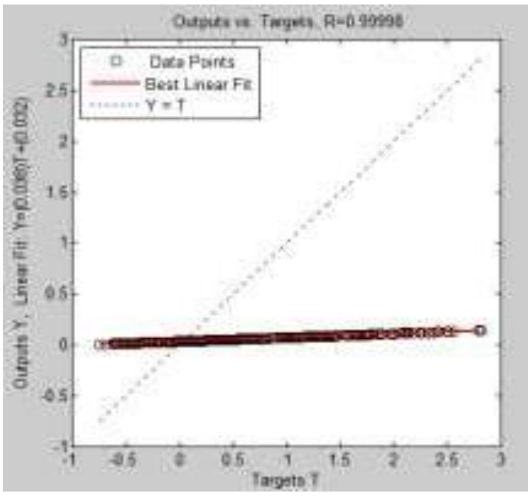

Fig.6. Regression Plot For Male and Female

TABLE 1. EFFECT OF USING DWT AT CLASSIFICATION ON NOISY SIGNALS

Noise effect					
DWT Level J	Training network	Transfer Function	mse	SNR	Rate[%]
0	Trainb	Tansig	0.009	20	100
0	Trainb	Tansig	0.009	-6.9	50.2
1	Trainb	Tansig	0.009	-6.9	56.2
2	Trainb	Tansig	0.009	-6.9	58.3
3	Trainb	Tansig	0.009	-6.9	52.1
8	Trainb	Tansig	0.009	-6.9	63.2
15	Trainb	Tansig	0.009	-6.9	68.3

V. CONCLUSIONS

In this paper, Wavelet Transform based feature extraction method is investigated. The introduced system depends on two steps gender discrimination and features extracting due to its better accuracy. The system works with excellent capability of features tracking even with -6.9 Db SNR., which is suitable for non-Stationary signal. Text-dependent system is used, so that the system can be applied in password or PINs identification in any security system, Banks, Hotels rooms, or other companies. One thousand speech signals are tested. The results show excellences performance with 98% identification.

VI. REFERENCES

[1]. K.Drqrouq, T.Abu Hilal, M Sharif,s.and A-Al Qawasmi,"Speaker identification using wavelet and neural network", IEEE Transctions on computers,July2009

[2]. M.a.Al-Aloaoui, „A new Weighted Generalized Algorithm for Pattern Recognition." IEEE Transactions on computers, vol. C- 5,no 10,October 1977.

[3]. Gabor, „Theory of communication" Journal of IEEE, 1993.

[4]. J.M.Naik, L.P.Nestsh and G.R. Doddingont , „speaker Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing , Glasgow, Scotland, May 1989, pages 524-527,

[5]. B Abdel-Rahman Al-Qawasmi and Khaled Daqrouq; „Discrete Wavelet Transform with Enhancement Filter for ECG signal".

[6]. M.AMAL-Alaoui Some Applications of Pattern Recognition , Ph.D. Thesis electric engineering department, Georgia Institute of December, 1974

[7]. An Efficient Feature Selection Method For Speaker Recognition,Hanwu sun, Bin Ma and Haizhou Li,Insititute of infocomm research.

[8]. Institute for Infocomm Research agency for Science, Technology and research, Singapore

[9]. T Matsui and S. Furui. Concatenated phoneme models for IEEE Proceedings of the 1993 International conferences acoustic speech and signal and speech signal processing , April 1993.

[10]. Chakroborty, S., Roy, A. and Saha, G., "Improved Independent speaker identification by combining MFCC with vidence from Flipped Filter Banks",

[11]. T. Matsui and S. Furui. Comparison of text-independent speaker recognition methods using VQ-distortion and discreet continuous HMM, IN IEEE Procedeeing of 1999.

[12]. Speech Recognition using Neural Networks,Joe Tebelsiks,may 1995,cmu-cs-95-142.

[13]. Wavelet tutorial by R.B.Polikar,volume III AND VOL.IV,on computers,

[14]. Liu C. H., Chen O. T. C. A Text independence speaker Identification system using PARCOR and AR model Vol 3- 335-336,2002.

[15]. The GABOR „Theory of Communication, Journel of I.E.E. 93- 97 PP ,429-19 . 1996

[16]. Introduction to speech recognition Kimberlee A. Kemble, program manager,IBM Coporation,20008.

[17]. Wavelet denoising and speech enhancement, by V.Balakrishanan

Cite this article as :

Dr. Vinodpuri Rampuri Gosavi, Dr. Manju D. Pawar, Dr. Ganesh S. Sable, "Speaker Verification using CWT DWT Transform and Neural Network", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 3, pp. 125-130, May-June 2022. Available at doi : https://doi.org/10.32628/IJSRST229317

Journal URL : https://ijsrst.com/IJSRST229317