The Use of Wavelets in Speaker Feature Tracking Identification System Using Neural Network

WAEL AL-SAWALMEH*, KHALED DAQROUQ*, ABDEL-RAHMAN AL-QAWASMI*,
AND TAREQ ABU HILAL**

*Department of Communication and Electronics Engineering
Philadelphia University
P.O. Box 1, Amman 19392
JORDAN
narin912007@yahoo.com haleddaq@yahoo.com qawasmi@philadelphia.edu.jo
http://www.philadelphia.edu.jo
**Foundation Program
Dhofar University
OMAN
tr_helal@yahoo.com

Abstract— Continuous and Discrete Wavelet Transform (WT) are used to create text-dependent robust to noise speaker recognition system. In this paper we investigate the accuracy of identification the speaker identity in non- stationary signals. Three methods are used to extract the essential speaker features based on Continuous, Discrete Wavelet Transform and Power Spectrum Density (PSD). To have better identification rate, two types of Neural Networks (NNT) are studied: The first is Feed Forward Back Propagation Neural Network (FFBNN) and the second is perceptron. Up to 98.44% identification rate is achieved. The presented system depends on the multi-stage features extracting due to its better accuracy. The multistage features tracking based system shows good capability of features tracking for tested signals with SNR equals to -9 dB using Wavelet Transform, which is suitable for non-stationary signal.

Key-words — Speaker identification; Continuous and discrete wavelet transform; Linear prediction coefficient; and text-dependent.

1 Introduction

Speaker recognition or voice classification is the task of recognizing people from their voices. Such systems extract features from speech signal, process them and use them to recognize the person from the voice. There is a difference between speaker recognition (recognizing who is speaking) and speech recognition (recognizing what is being said).
Speech recognition systems are now available commercially for a variety of purposes, such as voice dictation on computers and voice dialing on mobile phones. Automated speaker Recognition (ASR) systems have immediate advantages in any application requiring high degree of security as required in the banking sector and military, among others. Text dependent applications for the task of speaker recognition typically out-per ASR form their text independent counter parts due to the more simple application of the ASR task. Text-dependent refers to the speaker having to say a set utterance for identification, as opposed to text- independent approaches which are largely invariant to the type of utterance [1-3].

Over last four decays many solutions of speaker recognition have been appeared in literatures [10-12].

The Al-Alaoui algorithm for pattern classification [4-7] was motivated by Patterson and Womack's [8] and Wee's [9] proofs that the Mean Square Error (MSE) solution of the pattern classification solution gives a minimum mean-square-error approximation to Bayes' discrimination, weighted by the probability density function of the sample. All audio techniques start by converting the raw speech signal into a sequence of acoustic feature vectors 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 Cepstral Coefficients (MFCC) [13,14], Linear Prediction Cepstral Coefficients (LPCC) [15-17], and Perceptual Linear Prediction Cepstral (PLPC) Coefficients. All these features are based on the spectral information derived from a short time windowed segment of speech signals

One of the most common short-term spectral measurements currently used are Linear Predictive Coding (LPC) derived cepstral 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 provides a more stable representation from one repetition to another of a particular speaker's utterances. As for the regression coefficients, typically the first- and second-order coefficients are extracted at every frame period to represent the spectral dynamics. These coefficients are derivatives of the time functions 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 utterance is represented by a sequence of feature vectors, generally short-term spectral feature vectors. The time axes of the input utterance and each reference template or reference model of the registered speakers are aligned using a dynamic time warping Discrete wavelet Transform (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 [27].

One of the most successful text-independent recognition methods is based on Vector Quantization (VQ). In this method, VQ codebook s consisting 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 of 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 parameters over the long term can be represented by stochastic Markovian transitions between states. Therefore, methods using an ergodic HMM, where all possible transitions between states are allowed, have been proposed. Speech segments are classified into one of the broad phonetic categories corresponding to the HMM states. After the classification, appropriate features are selected.

It has been shown that a continuous ergodic HMM method is far superior to a discrete ergodic HMM method and that a continuous ergodic HMM method is as robust as a VQ-based method when enough training data is available [25].

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].

2 Vocal Tract Model

When a person speaks the lungs work like a power supply of the speech generating system. The glottis supplies the input with the certain pitch frequency (F_0). The vocal tract, which contains the pharynx and the mouth and nose cavities, works like a musical instrument to generate a sound. In fact, different vocal tract character or shape would generate a different sound (wave). To form distinct vocal tract shapes, the mouth cavity plays the important role. Nasal cavity is often included in the vocal tract system. The nasal cavity and the mouth cavity are connected in parallel. The vocal tract model is shown in Figure. 1.

The glottal pulse produced by the glottis is used to generate vowels or sounds. And the noise-like signal is used to produce consonants or unvoiced sounds Pitch frequency F0 = (1/T0) varies in different people. A child's pitch frequency is as high as 400 Hz. Adult male's pitch frequency is as low as 100 Hz. Adult female's pitch frequency varies from 200 Hz to 300 Hz [23].

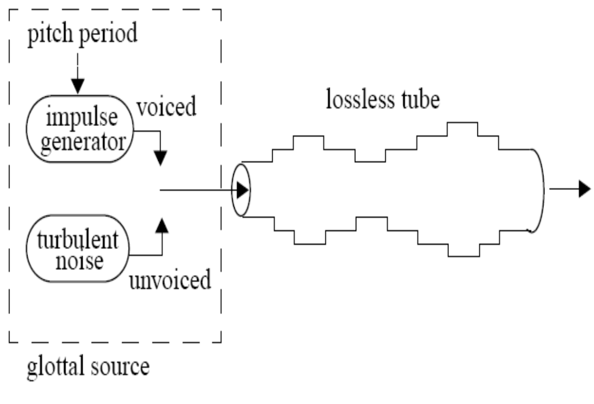


Figure 1: Vocal tract model

In this research significant in term of using the Wavelet Transform based recognition system is presented. This system is divided into two main blocks; features extracting and identification. The advantage of the system is Wavelet transform using. The speech signal is given to three stages of features extracting and identification based on mainly

Wavelet Transform. This transform that depends on convolution with wavelet functions, can track the very quick variation in frequency changing. That what exactly happens in non-stationary signal such as speech signal? The paper contains Wavelet Transform and NNT descriptions but Discrete Wavelet Transform is described in the method section. Presented system performance is described in results and discussion section, and finally, conclusion is introduced.

3 Continuous Wavelet Transform

The Fourier transform is very widely used tool for many mathematical or scientific applications, but it is well suited only to the study of stationary signals where all frequencies have an infinite coherence time. The Fourier analysis brings only global information which is not enough to detect compact patterns. Gabor introduced a local Fourier analysis, taking into account a sliding window [18], leading to a time frequency-analysis. This methodology is only applicable to problems where the coherence time is independent of the frequency. This is the case for instance for singing signals which have their coherence time determined by the geometry of the oral cavity. Morlet has introduced the Wavelet Transform [21] in order to have a coherence time proportional to the period.

The Morlet-Grossmann definition of the continuous wavelet transform [20] for a one dimensional 1D signal $f(x) \in L^2(R)$ is:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi^* \left(\frac{x-b}{a} \right) dx \tag{1}$$

where z* denotes the complex conjugate of z, $\psi^*(x)$ is the analyzing wavelet (Figure.2), a (>0) is the scale parameter and b is the position parameter.

The transform is characterized by the following three properties:

1. It is a linear transformation,
2. It is covariant under translations:

$$f(x) \rightarrow f(x-u) \quad W(a,b) \rightarrow W(a,b-u) \tag{2}$$

3. It is covariant under dilations:

$$f(x) \rightarrow f(sx) \quad W(a,b) \rightarrow s^{-\frac{1}{2}} W(sa,sb) \tag{3}$$

The last property makes the wavelet transform very suitable for analyzing hierarchical structures. It is like a mathematical microscope with properties that do not depend on the magnification.

In Fourier space, we have:

$$\hat{W}(a,v) = \sqrt{a} f(v) \psi^*(a,v) \tag{4}$$

When the scale a varies, the filter $\psi^*(av)$ is only reduced or dilated while keeping the same pattern.

Now consider a function $W(a,b)$ which is the wavelet transform of a given function *f(x)*. It has been shown that f(x) can be restored using the formula:

$$f(x) = \frac{1}{C_\chi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{a}} W(a,b) \chi \left(\frac{x-b}{a} \right) \frac{da.db}{a^2}$$

where:

$$C_x = \int_0^{+\infty} \frac{\psi^*(v) \hat{\chi}(v)}{v} dv = \int_{-\infty}^{0} \frac{\psi^* \hat{\chi}(v)}{v} dv \tag{5}$$

Generally $\chi(x) = \psi(x)$, but other choices can enhance certain features for some applications.

The reconstruction is only available if C_χ is defined (admissibility condition). In the case of $\chi(x) = \psi(x)$, this condition implies $\hat{\psi}(0) = 0$, i.e. the mean of the wavelet function is 0.

The wavelet defined by Morlet [19] is a complex wavelet which can be decomposed in two parts, one for the real part, and the other for the imaginary part.

$$\begin{aligned} g_r(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cos(2\pi v_0 x) \\ g_i(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sin(2\pi v_0 x) \end{aligned} \tag{6}$$

where v_0 is a constant. The admissibility condition is verified only if $v_0 > 0.8$.

Figure 1 shows these two functions.

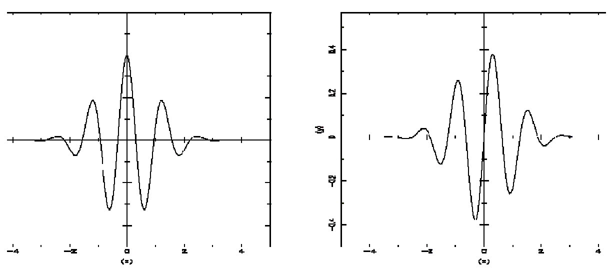


Figure 2: Morlet's wavelet: real part at left and imaginary part at right.

4 Proposed Method

The proposed system based on two main blocks. In the first block feature extracting is accomplished by Wavelet Transform and PSD. But the second block presents identification process via verification by NNT.

The first stage of this method is to decompose the speech signal into Continuous Wavelet Transform sub-signals of given scale that must based on speaker own feature frequency depending on its anatomical structure of his own vocal tract and other working parts through speaking process (Figure. 1). The continuous wavelet transform scale determination is very challenging problem because of its non-stationary nature contained in speech signals; therefore we use experimental scale determination by studying huge data base of about 1000 speech signals. This assists greatly in finding out one scale matching all people. To extract sharp and more concentrated features, we use PSD (figure. 7). Now if these features match any of our models stored in the system, system will go on to next stage, if not, system will cancel the trail. To take matching decision two NNT are studied: FFBNN and Perceptron.

The second stage of feature tracking method is discrete wavelet transform feature tracking. In this stage the speech signal is decomposed into discrete wavelet transform sub-signals ($d_1, d_2, \ldots, d_J$), where each of these sub-signals is generated by Mallat's algorithm of particular level $1,2,\ldots,J$. This accomplished by convolution the signal with mother wavelet function to achieve high pass sub-signals of speech signal. And s_J , which is accomplished by convolution with father wavelet function (scaled version) to achieve low pass sub-signals of the speech signal.

Now speaker vocal tract frequency which is known as low frequency is contained in d_5, where $J = 5$, which is determined imperially by studying more than 1000 speech signals (Figure. 8). Also, PSD is used. In this stage, the system can cancel the trail if no match is achieved, too.

The decision is taken (in identification block) by the system by NNT.

The third stage of the tracking method is accomplished by dividing the signals into windows of equal duration and calculating continuous wavelet transform of each window. This helps to extract the small speech parts feature by using PSD.

This stage creates more accurate features in different parts of a speech signal. To each window NNT is applied and each window should pass.

This identification system advantage is that it works via three features extraction stages. For each stage identification block is applied by NNT classification system using Matlab.

Artificial neural networks have advanced in leaps and bounds since their invention in 1943 and their first implementation to tackle real world problems solutions in 1958.

The history of neural networks, a general description of neural networks, the different types of architectures and the networks associated with them as well as their applications is presented in many references.

The latest developments in the research of neural networks are providing society with new and exciting methods of dealing with complicated problems and tedious tasks. It can be concluded that the future of artificial neural networks and artificial intelligence looks very promising [28].

Feed-Forward (FF) Back propagation Network that is applied in our research is implemented by using Matlab function new FF, which consists of N layers using the dot product weight function, net sum net input function, and the specified transfer function.

The first layer has weights coming from the input. Each subsequent layer has a weight coming from the previous layer. All layers have biases. The last layer is the network output. Each layer's weights and biases are initialized with Matlab command.

Adaption can be done with trains (Backdrop network training function) or (Backdrop weight/bias learning function), which updates weights with the specified learning function. Training is done with the specified training function. Performance is measured according to the specified performance function.

Back-propagation is the most popularly used method for training multi-layer feed-forward networks. For most networks, the learning performance is based on error function, which is then minimized with respect to the weights and bias. Therefore, we can evaluate the derivative of the error with respect to weights, and these derivatives can then be used to find the weights that minimize the error function, by either using the popular gradient descent or other optimization methods. The algorithm for evaluating the derivative of the error function is known as back-propagation, because it propagates the errors backward through the network [28].

Perceptron is common a pattern recognition machine, based on an analogy to the human nervous system, capable of learning in term of a feedback system which reinforces correct answers and discourages wrong ones. It is a type of artificial neural network was created in 1957 at the Cornell Aeronautical Laboratory by Frank Rosenblatt. It can be seen as the simplest type of feed-forward neural network and a linear classifier.

The Perceptron is a binary classifier that make its input x which is real-valued vector an output value *f(x)* (a single binary value) through the matrix.

$$f(x = \begin{cases} 1 & if \quad \omega..x + b > 0 \\ 0 & \quad else \end{cases} \tag{7}$$

where ω is a vector of real-valued weights and $\omega.x$ is the dot product (which computes a weighted sum). b is the 'bias'.

The value of *f(x)* (0 or 1) is used to classify x as either a positive or a negative instance, in the case of a binary classification problem

Since the inputs are fed directly to the output unit via the weighted connections, the perceptron can be considered the simplest kind of feed-forward neural network.

The learning algorithm is the same across all neurons; therefore everything that follows is applied to a single neuron in isolation. We define *x(j)* as the *j-th* item in the n-dimensional input vector, *w(j)* as the *j-th* item in the weight vector, f(x) denotes the output from the neuron when presented with input x and α is a constant where $0 < \alpha \leq 1$ is learning rate.

Further, assume for convenience that the bias term b is zero. This is not a restriction since an extra dimension n + 1 can be added to the input vectors x with *x(n + 1) = 1*, in which case *w(n + 1)* replaces the bias term (Figure.3).

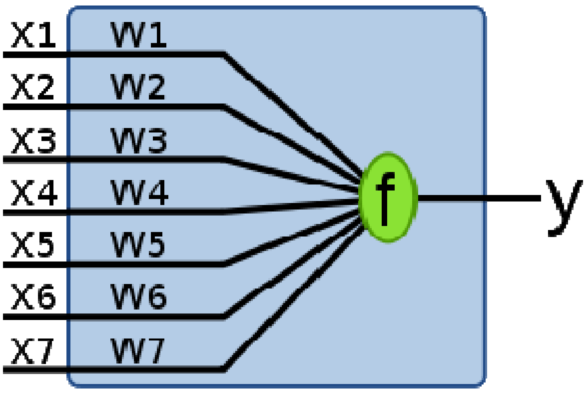


Figure 3: Perceptron network architecture

The suitable weights are applied to the inputs, and the resulting weighted sum passed to a function which produces the output y.

Learning is process of the weight vector being updated for multiple iterations over all training examples.

Let $D_m = \{(x_1, y_1), \cdots, (x_m, y_m)\}$ denote a training set of m training examples.

Iteration of the weight vector is updated as follows:

-For each *(x,y)* pair in $D_m = \{(x_1, y_1), \cdots, (x_m, y_m)\}$ and

$w(j) := w(j) + \alpha(y - f(x))x(j) \quad (j = 1, \cdots, n)$

Note that this means that a change in the weight vector will only take place for a given training example *(x,y)* if its output *f(x)* is different from the desired output y.

The initialization of w is usually performed simply by setting *w(j): = 0* for all elements *w(j)*.

The training set Dm is said to be linearly separable if there exists a positive constant γ and a weight vector w such that $y_i.(\langle w, x_i \rangle + b) > \gamma$ for all *i*.

5 Results and Discussion

The testing speech signals data base consist of about *1000* speech signal that were recorded via PC-sound card, with spectral frequency *4000* Hz and sampling frequency *16000 Hz* over about *2 sec*. time duration. Each speaker recorded "Besme Allah Alrahman Alraheem" Arabic word that means "in the name of God" in English. Each utterance of one "Besme Allah Alrahman Alraheem" words was recorded 10 times by the speaker.

From above speech signal discretion, we can notice that presented recognition system is text-dependent system, because prompts are common across all speakers that can share secrets (passwords or PINs). In order to create multi-factor authentication scenario, the speaker in each trail is compared to all models stored in database.

Figure.4 illustrates the 120 coefficients of PSD of 8 speech signals for two persons, where (• blue) coefficients present speaker1 and (o red) coefficients present speaker2. Unfortunately, coefficients from 20 to 120 are totally overlapped. In Figure.5 and 6 illustrate the effect of Continuous Wavelet Transform (CWT) on PSD separating of the two speakers. The utilities of separating depend on the J level which is related to special band pass of frequency achieved by wavelet filters.

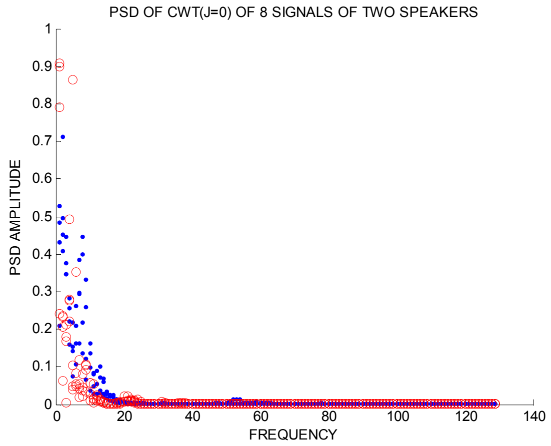


Figure 4: 120 PSD coefficients of eight speech signals, each four signals belong to one speaker. (• Blue) coefficients present speaker1 but (o Red) coefficients present speaker2.

.

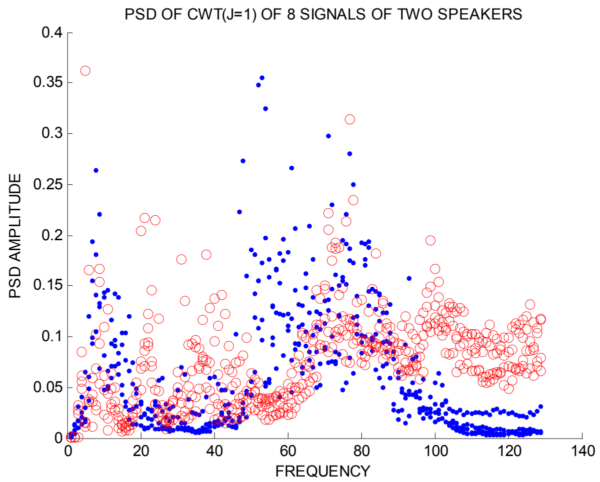


Figure 5: 120 PSD coefficients of CWT level J=1 of eight speech signals, each four signals belong to one speaker. (• Blue) coefficients present speaker1 but (o Red) coefficients present speaker2

For speech signals identification via verification two Neural Networks are studied: FFBNN and Perceptron

Table 1 shows the results of Network FFBNN by newff Matlab function. We can notice the ability of identification based on this algorithm. The first part of the table shows the results FFBNN using network back propagation network training function. In the second part of table we train a network with weight and bias learning rule with batch updates. The weights and biases are updated at the end of an entire pass through the input data (batch training). Better results were achieved in first part.

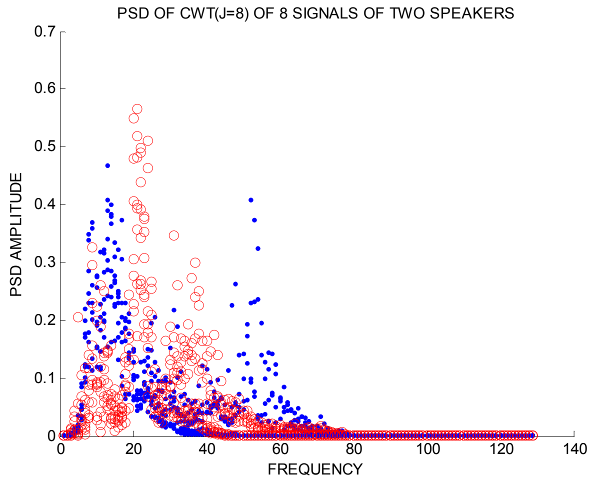


Fig.6:Same PSD results of CWT of *J=8* of same speech signals used in Fig.4,5, and 6

Part three is presented to study the system noise robustness, to same 8 signals. Strong sinusoidal noise was added. Signal-to-noise ratio (SNR) was (-9 dB). In this case PSD without CWT is totally unsuccessful where recognition rate was less than 50%. Vice versa CWT is more robust to noise in separating PSD of very noised signals. This effect can be seen at spectrogram in figure 11, where the use of CWT concentrates the signal energy.

For same purpose we create a perceptron NNT of input matrix of PSD (*P=[t1 ; t1]*), t1 is PSD of CWT of user one and binary target (*T=[0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1]*). We train the network to classify t1 as 0 and same t1 as 1 and simulate other signal t2 of different speaker, and plot perceptron input/target vectors in figure8. 1 is presented as (+). We notice that signal is perfect classified where 0 overlap ones. But test input t2 is separated by 0 red color.

Figure 9 and 10 illustrate the plot a classification line on a perceptron vector plot. In figure 8, P =[t1 ,t2] but P test is [t1 ,t1]. The classification is clear that perceptron plot shows, after simulating with P test (red 0) that is classified as 0 like t1 in P. the two classes are separated by the line. In figure 10 similar experiment is illustrated but with less accuracy.

6 Conclusions

The effect of Wavelet Transform on speaker feature extracting is studied. The introduced system in this paper depends on multi-stage features extracting due to its better accuracy. The system works with capability of features tracking even with -9 dB SNR. This is accomplished because of multistage features tracking based system using Wavelet Transform, which is suitable for non-stationary signal. Text -dependant system is used, so that the system can be applied in password or PINs identification in any security system, Banks, Hotel rooms, or other companies. Two NNT methods are investigated FFBNT perceptron NNT. One thousand speech signals are tested. The results show excellent performance with more than 70% up to 98.44% identification rate.

References

[1] Simon Lucey, Audio-visual Speech Processing PhD Thesis, School of Electrical & Electronic Systems Engineering, Queensland University of Technology, April 2002.

[2] http://www.itl.nist.gov/div893/biometrics/Biometris from the movies. pdf, National Institute of Standards and Technology, Biometric Authentication Technology: From the Movies to Your Desktop by Fernando L. Podio1 and Jeffrey S. Dunn2,

[3] Elisabeth Zetterholm, Voice Imitation. A Phonetic Study of Perceptual Illusions and Acoustic Success. Phd thesis, Lund University. (2003).

[4] M. A. Al-Alaoui, Some Applications of Generalized Inverse to Pattern Recognition, Ph.D. Thesis, Electrical Engineering Department, Georgia Institute of Technology, December, 1974.

[5]] M. A. Al-Alaoui, “A New Weighted Generalized Inverse Algorithm for Pattern Recognition.” IEEE Transactions on Computers, Vol. C-26, No. 10, pp. 1009-1017, October 1977.

[6] M. A. Al-Alaoui, “Application of Constrained Generalized Inverse to Pattern Recognition.” Pattern Recognition, Pergamon Press, Vol. 8, pp. 277-281, 1976.

[7] M. A. Al-Alaoui, J. El Achkar, M. Hijazi, T. Zeineddine and M. Khuri, "Application of Artificial Neural Networks to QRS Detection and LVH Diagnosis"; Proceedings of ICECS'95, pp. 377-380, Amman-Jordan, 17-21 December 1995.

[8] [8] J. D. Ptterson and B. F. Womack, “An Adaptive Pattern Classification System,” IEEE Transactions Syst. Man. Cybern. , Vol. SSC-2, pp. 62-67, August 1966.

[9] W. G. Wee, “Generalized Inverse Approach to Adaptive Multiclass Pattern Classification,” IEEE Transactions on Computers, Vol. C-17, pp. 1157-1164,
December 1968.

[10] Special Issue on Speaker Recognition, Digital Signal Processing, vol. 10, January 2000.

[11] R. Teunen, B. Shahshahani, and L. Heck, ``A Model-based Transformational Approach to Robust Speaker Recognition,'' ICSLP October 2000.

[12] A. Higgins, L. Bahler, and J. Porter, ``Speaker Verification using Randomized Phrase Prompting,'' Digital Signal Processing, vol. 1, pp. 89-106,1991

[13] Chakroborty, S., Roy, A. and Saha, G., “Improved Closed set Text-Independent Speaker Identification by Combining MFCC with vidence from Flipped Filter Banks”. International Journal of Signal Processing, Vol. 4, No. 2, Page(s):114-122, 2007.

[14] Sandipan Chakroborty* and Goutam Saha, Improved Text-Independent Speaker Identification using Fused MFCC & IMFCC Feature Sets based on Gaussian Filter. International Journal of Signal Processing 5;1 © www.waset.org Winter 2009.

[15] Sanderson S. Automatic Person Verification Using Speech and Face Information. PhD thesis. Griffith University. 2002.

[16] Petry A. and Barone D. A. C. Fractal Dimension Applied to Speaker Identification. ICASSP (Salt Lake City). May 7-11. 405-408, 2001.

[17] Liu C. H., Chen O. T. C. A Text-Independent Speaker Identification System Using PARCOR and AR Model. MWSCAS. Vol 3, 332-335. 2002.

[18] D. Gabor, ``Theory of communication", Journal of I.E.E. 93 pp 429-441, 1946.

[19] P. Goupillaud, A. Grossmann, J. Morlet, ``Cycle-octave and related transforms in seismic signal analysis", Geoexploration, 23, 85-102, 1984-1985.

[20] A. Grossmann and J. Morlet, ``Decomposition of Hardy functions into square integrable wavelets of constant shape", SIAM J. Math. Anal, Vol. 15, pp 723-736, 1984.

[21] Meyer, Wavelets, Ed. J.M. Combes et al., Springer Verlag, Berlin, p. 21, 1989.

[22] Abdel-Rahman Al-Qawasmi and Khaled Daqrouq "Discrete Wavelet Transform with Enhancement Filter for ECG Signal", WSEAS transaction on Biology and Biomedicine" 2009

[23] http://www.dtic.upf.edu/~xserra/cursos/TDP/referencies/Park-LPC-tutorial.pdf

[24] T. Matsui and S. Furui. Concatenated phoneme models for text-variable speaker recognition. IEEE Proceedings of the 1993 International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, Minnesota, April 1993, pages 391--394.

[25] T. Matsui and S. Furui. Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs. In

IEEE Proceedings of the 1993 International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, Minnesota, April 1993,], pages 157--160.

[26] C. Griffin, T. Matsui, and S. Furui. Distance measures for text-independent speaker recognition based on MAR model. IEEE Proceedings of the 1994 International Conference on Acoustics, Speech, and Signal Processing, Adelaide, Australia, April 1994, pages 309--312.

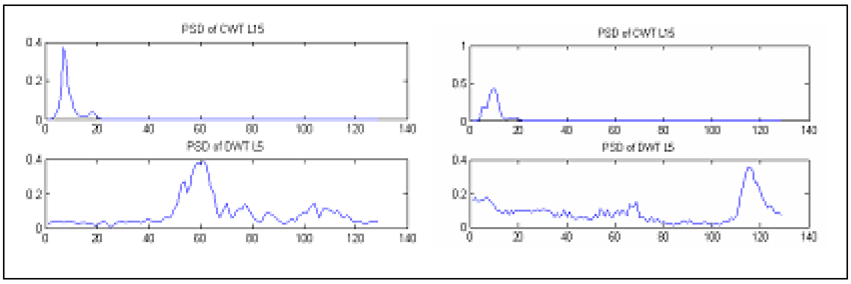
[27] J. M. Naik, L. P. Netsch, and G. R. Doddington. Speaker verification over long distance telephone lines. In IEEE Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing, Glasgow, Scotland, May 1989, pages 524--527.

[28] A. El Shahat, Neural Unit for PM Synchronous Machine Performance Improvement Used For Renewable Energy, GCREEDER 2009, Amman-Jordan, March 31st – April 2nd 2009.

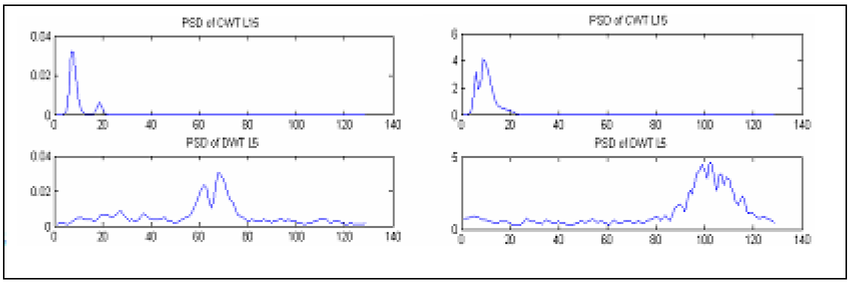
[29] Answers.com, Search Engine of unanswered questions.

[30] Freund, Y. and Schapire, R. E. 1998. Large margin classification using the perceptron algorithm. In Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT' 98). ACM Press.

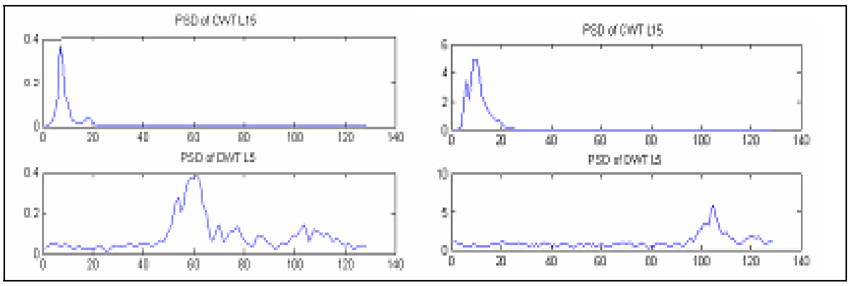
[31] Freund, Y. and Schapire, R. E. 1999. Large margin classification using the perceptron algorithm. In Machine Learning 37(3):277-296, 1999.



Trail1



Trail2



Trail3

Fig.7: Three tails of the first stage (Power Spectrum Density of Continuous Wavelet Transform level 15) at the top and the second stage (Power Spectrum Density of Discrete Wavelet Transform level 5) at bottom, at left is speaker1, and at right is speaker2

Trainlm Training Network					
CWT Level J	*Training network*	*Transfer Function*	*MSE*	*Gradient*	*Rate[%]*
0	Trainlm	Tansig	0.0078	$1.8*10^{-12}$	63.1
1	Trainlm	Tansig	0.0078	$5.7*10^{-9}$	73.1
2	Trainlm	Tansig	0.0078	$1.4*10^{-14}$	73.1
3	Trainlm	Tansig	0.0078	$1.4*10^{-13}$	70.6
4	Trainlm	Tansig	0.0078	$5.7*10^{-14}$	75.3
8	Trainlm	Tansig	0.0078	$3.3*10^{-9}$	73.7
15	Trainlm	Tansig	0.0078	$1.* 10^{-9}$	74.6
Trainb Training Network					
CWT Level J	*Training network*	*Transfer Function*	*MSE*	*Gradient*	*Rate[%]*
0	Trainb	Tansig	0.009	Not shown	59.1
1	Trainb	Tansig	0.009	Not shown	72.1
2	Trainb	Tansig	0.009	Not shown	67.1
3	Trainb	Tansig	0.009	Not shown	70.5
4	Trainb	Tansig	0.009	Not shown	70.6
8	Trainb	Tansig	0.009	Not shown	70.6
15	Trainb	Tansig	0.009	Not shown	72.2
Noise effect					
CWT Level J	*Training network*	*Transfer Function*	*MSE*	*SNR*	*Rate[%]*
0	Trainb	Tansig	0.009	20	98.4
0	Trainb	Tansig	0.009	-6.9	42.2 to 49
1	Trainb	Tansig	0.009	-6.9	55.1
2	Trainb	Tansig	0.009	-6.9	58
3	Trainb	Tansig	0.009	-6.9	51.2
8	Trainb	Tansig	0.009	-6.9	61.1
15	Trainb	Tansig	0.009	-6.9	58.2

Table 1: FFBNN results

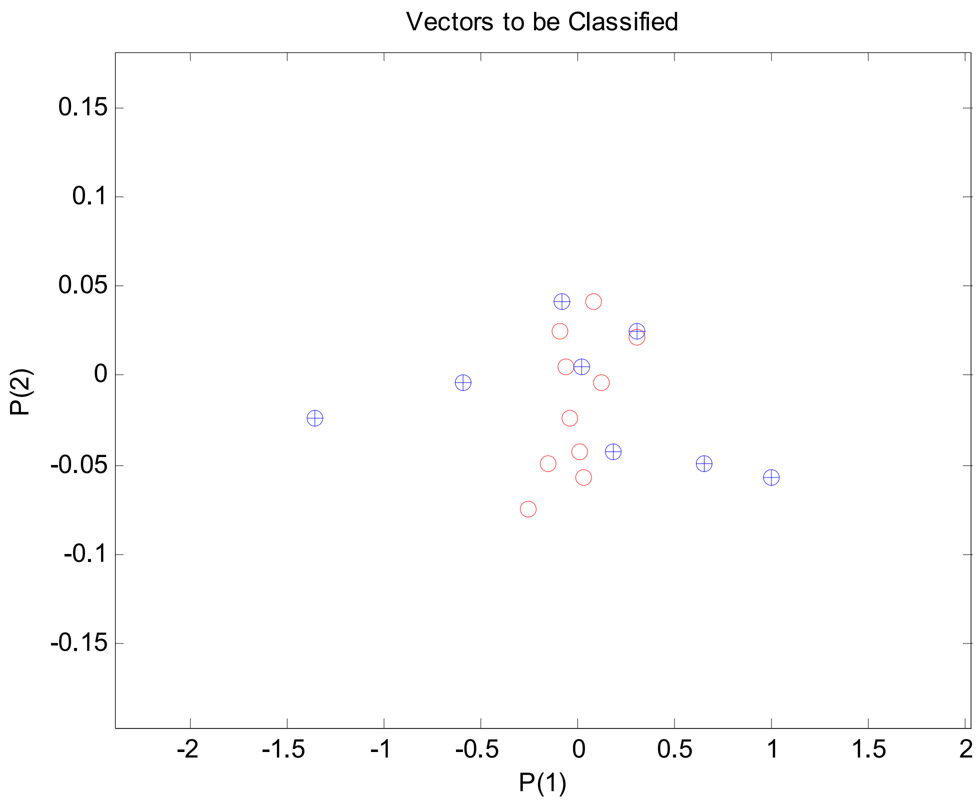


Fig.8: Plot of perceptron input/target vectors

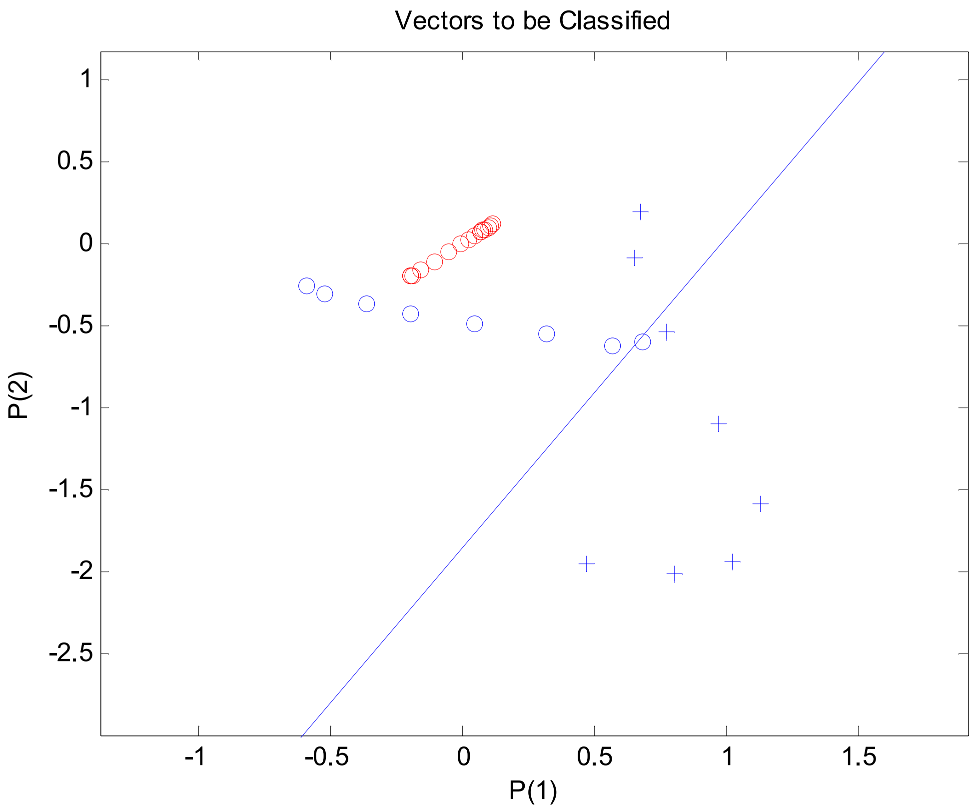


Fig.9: Illustrates the plot a classification line on a perceptron vector plot

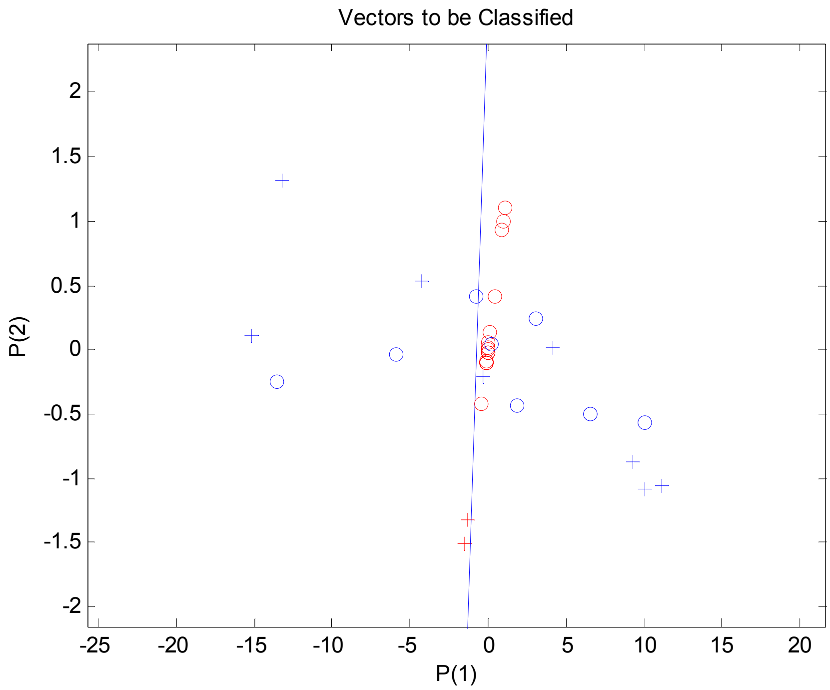
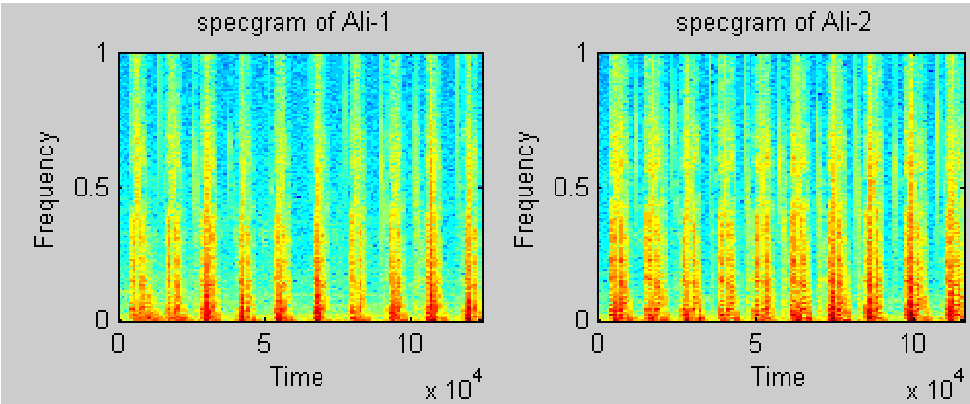
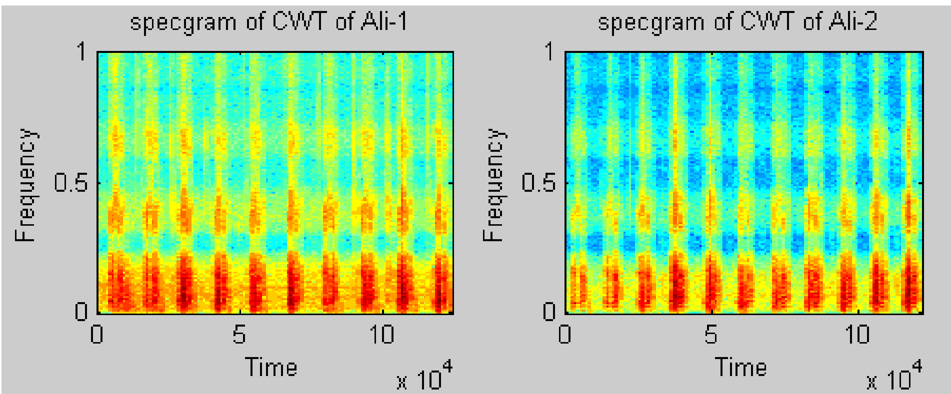


Fig.10: Illustrates the plot a classification line on a perceptron vector plot



a.



b.

Fig.11: Effect of CWT on the spectrogram density. a. Ali two utterances without CWT and b. Ali two utterances with CWT.