World of Computer Science and Information Technology Journal (WCSIT)
ISSN: 2221-0741
Vol. 2, No. 2, 68-73, 2012

A Novel Study of Biometric Speaker Identification Using Neural Networks and Multi-Level Wavelet Decomposition

Aryaf Abdullah Aladwan
Computer Engineering dept.
Faculty of Engineering Technology
Al-Balqa' Applied University
Amman , Jordan

Rufaida Muhammad Shamroukh
Computer Engineering dept.
Faculty of Engineering Technology
Al-Balqa' Applied University
Amman , Jordan

Ana'am Abdullah Aladwan
Information systems & technology dept.
University of Banking and Financial sciences
Amman , Jordan

Abstract — Researchers in voice and speaker recognition systems has been entered a new stage. The overall researches concern on trials to enhance the accuracy and precession of the developed system techniques, especially in intelligent systems. The use of Digital Signal Processing (DSP) with cooperation of Artificial Intelligence (AI) is common in such researches. But the main inertia in that is to developing the algorithm in trial and error in most cases. This research aims to find the hot spot points in merging specific techniques of DSP with AI. Neural networks based speaker recognition was been developed in order to test the results of the proposed algorithm and record the study results. Multi-level decomposition of wavelet transformation is adopted to extract the features of the speaker person. The feature extraction using wavelet transformation is studied and this paper determines the best level and condition of applying that technique.

Keywords-Wavelet Transform; Multi-Level Decomposition; Voice; Neural Networks; Speaker Identification; Biometrics.

I. INTRODUCTION

Speaker verification deals with determining the identity of a given speaker using a predefined set of samples. Different recognizers could be used for speaker identification (i.e. neural networks, genetic algorithms, statistical approaches, etc) depending on different set of features extracted from the person voice such as LPC (Linear predictive coefficient), wavelet transformation (discrete and continues), DCT discrete cosine transform, etc. The main steps of voice recognition starts with preprocessing the voice signal by perform sampling and quantization; this depends on the voice acquisition tool that is being used; then performs feature extraction after wavelet transformation. Finally, the extracted features are fed to a pattern recognition phase (classifier).

This field is still under intensive study at which the appropriate feature set that contains the best unique characteristic of each voice need to be investigated in addition to the appropriate classifier for each feature set.

The voice biometrics has a main place in computer systems and access controls. Voice and speaker recognition has a role of protecting the user's identity in addition to the computerized data. Such systems have become increasingly difficult. The main concept of security is authentication identifying or verifying.

The identity authentication could be done in three ways [1]:

1. Something the user knows. i.e. password
2. Something the user has. i.e. RFID
3. Something the user is. And this is so called Biometrics.

The biometrics is the concept of measuring unique features of human depending on bio-analysis. Such as a fingerprint, voice, face, etc.

A biometric system can operate in two modes the verification mode where the system performs a one to one comparison of a captured biometric with a specific template stored in a biometric database in order to verify the

individuals. And the identification mode where the system performs a one-to-many comparison against a biometric database in attempt to establish the identity of an unknown individual.

The accumulated problems of the traditional methods of human authentication cause a major importance of intelligent methods (Biometrics). The shortcoming in these methods that the key or credit card can be stolen or lost and PIN number or password can be easily misused or forgotten, such shortcoming will not be there in the biometric authentication.

One of the widely used systems is the voice recognition technique. Since every human have a unique feature in his voice, it is useful to discriminate between two persons using their own voices. The idea of voice recognition, which is different from speech recognition, is to verify the individual speaker against a stored voice pattern, not to understand what is being said. while speech recognition is concerned with understanding what is being said. In the field of voice recognition many techniques have been developed such as Hidden Markov Models, Neural network, Fuzzy logic and Genetic algorithms.

Human voice has two types of information high-level information and low level information. High-level information is values like dialect, an accent (the talking style and the subject manner of context).

Voice recognition deals with low-level information from the human speak voice, like pitch period, rhythm, tone, spectral magnitude, frequencies, and bandwidth of an individual's voice, this information taken as features. For voice recognition, another information attributes can be taken as features such like Mel-frequency Cepstrum Coefficients (MFCC) and Linear Predicative Cepstral Coefficient (LPCC). For robust voice recognition system, wavelet transform coefficients are used continuous or discrete. The concentration is on discrete wavelet.

The use of wavelet transform decomposition comes from the fact that, it has the most specs and recognizers of the speech and voice, including the person identifiers.

Many features can be extracted using wavelet decomposition, but the task of this paper is to determine the best analysis method that gets the best voice features.

There are two main factors should be selected in the design of wavelet based biometrics; number of levels, and the minimum set of coefficient extracted from level(s) that leads to better discrimination. This topic still under research since the most important thing is to keep the best recognition ability with minimum feature set to speed up the verification operation during searching huge voice dataset. This paper covers the two topics.

The bases functions that can be used in wavelet decomposition as the mother wavelet are including Haar wavelet, Daubechies wavelets, Coiflet1 wavelet, Symlet2 wavelet, Meyer wavelet, Morlet, Mexican Hat wavelet. In discrete case, the selection is being done between Harr wavelet and Daubechies wavelet. Hence, the Haar wavelet causes significant leakage of frequency components and is not well suited to spectral analysis of speech, whereas, the Daubechies family of wavelets has the advantage of having low spectral leakage and generally produces good results.

II. PROBLEM STATEMENT

The main contribution of this paper is to improve the matching process speed in the field of authentication system by suggesting minimum number of features that would not affect the system accuracy and study the effect of multi-level wavelet decomposition on speaker voice. The recognition system that is used to select the minimum feature set is feed forward Neural Network (Multi-Layer Perceptron) and learning vector quantization neural network. The suggested Neural Network will be trained with different sets of features extracted from different levels of discrete wavelet transform (DWT) then the trained recognition system will be tested using cross validation testing to determine the minimum feature set that is suitable for building voice recognition system.

The proposed approach consists of three phases, preprocessing phase, feature extraction phase, and recognition phase.

This research studies the features that extracted from different levels of discrete wavelet transformation and illustrate the effective of using different percent of each level instead of all features with feed forward and learning vector quantization neural network as classifier part. Then compare their recognition ability and decide the best level that is enough to give comparable result with respect to all feature set.

III. MULTI-LEVEL WAVELET DECOMPOSITION

The continuous wavelet transformation (CWT) is defined as the sum over all the time of a signal that multiplied by scaled and shifted wavelet function. The result

is a set of Wavelet coefficients, which are a function scale and position [2].

Dilation and translation of the Mother function, or analyzing wavelet Φ(x) defines an orthogonal basis, the wavelet basis is shown in equation -1:

$$\Phi_{(s,l)}(x) = 2^{\frac{-s}{2}} \Phi(2^{-s}x - l) \quad \dots (1)$$

The variables "s" and "l" are integers that scale and dilate the mother function Φ(x) to generate wavelets, such as a Daubechies wavelet family. The scale index s indicates the wavelet's width, and the location index "l" which gives its position. Notice that the mother functions are rescaled, or dilated by powers of two, and translated by integers. What makes wavelet bases especially interesting is the self-similarity caused by the scales and dilations. Once the mother functions are explained, everything about the basis will be clearer.

To span the domain of data at different resolutions, the wavelet analysis is used in the equation-2 (scaling) [2]:

$$W(x) = \sum_{k=-1}^{N-2} (-1)^k C_{k+1} \Phi(2x + k) \quad \dots (2)$$

Where W(x) is the scaling function for the mother function and Ck are the wavelet coefficients. The wavelet coefficients must satisfy quadratic and linear constraints of the form that shown in the equation-3 [2]:

$$\sum_{k=0}^{N-1} C_k = 2, \quad \sum_{k=0}^{N-1} C_k C_{k+2l} = 2\delta_{l,o} \quad \dots (3)$$

A special case of the wavelet transformation is the discrete wavelet transformation, which provides a compact representation of a signal in frequency and time. The discrete wavelet transform of a specified signal can be computed by passing the signal through series of low-pass and high-pass filters to analyze the frequencies. The outputs that generated are then down sampled by two, so the output is half of original signal size [3].

IV. METHODOLOGY

This system is developed to test the multi-levels wavelet decomposition of speaker voice validation and accuracy. The sound recognition system is developed using Multi-Layer perceptron neural network.

The main characteristic of any speaker recognition is the determination of the person who speaks. Speaker recognition system consists of several modules in addition to the classification engine. The proposed system consists of three main modules as shown in figure 1.

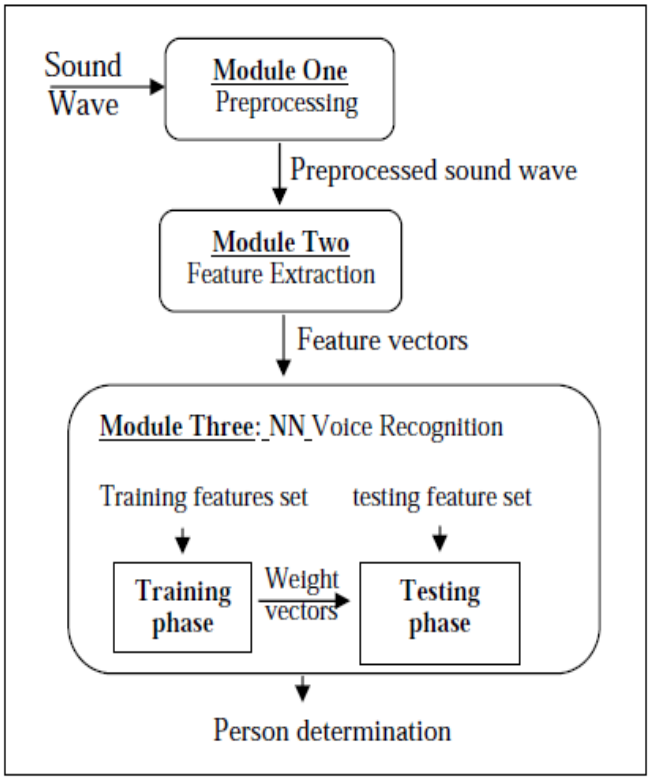

Figure 1. Basic Program Diagram of the Proposed System.

First, the human sounds are recorded and processed. The resulted sounds passed to the feature extraction module to extract features which represents the data set. This is the core of this paper, where it's done by multi-levels wavelet decomposition. Finally, the extracted features are fed to the recognition module, which is consists of two phase's neural network classifier; the training phase and testing phase. The trained system is used to recognize a person voice.

Noise removal is done using pre-filtration and DC-level removal. Figure 2 shows the signal after noise removing.

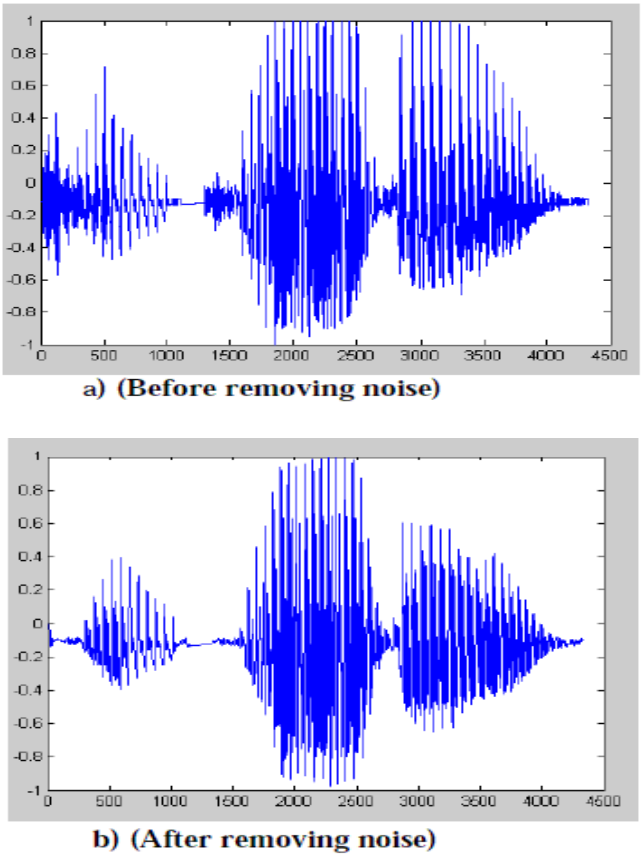

Figure 2. Noise Removal.

Once, the sound is being recorded and preprocessed, discrete wavelet transformation (DWT) is being applied on debauched level-1, level-2, to level-7. The key of using 7th level as the last one is that, the signal will be omitted after that level. Thus, each sound is ready to be passed to the neural network. After extracting the features (which represent the coefficients) of different wavelet transformation, the speaker recognition module is called.

To perform the recognition phases the MLP neural network is being used. Classification operation mainly consists of two phases (training phase and testing phase) as illustrated in figure-3.

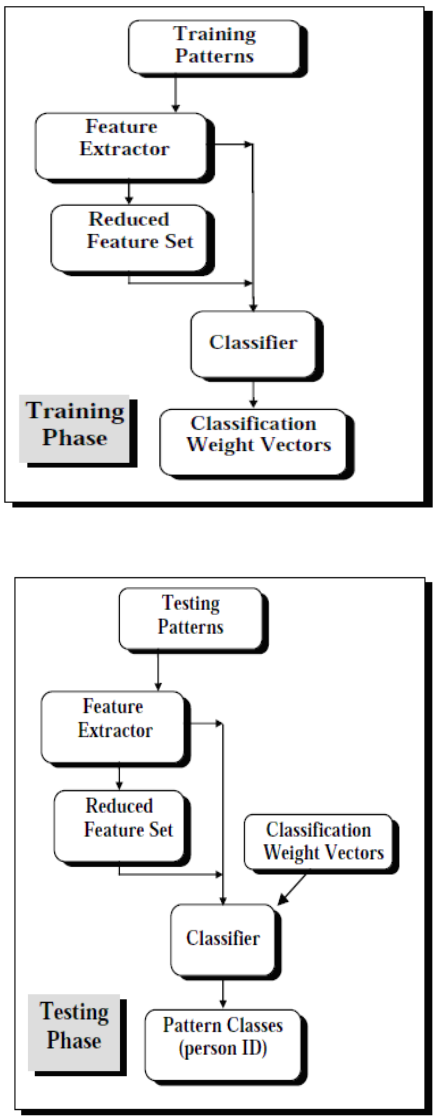

Figure 3. Training and testing phases.

The classifier, in training phase, is trained on a set of patterns (which represent set of DW coefficients extracted from different Wavelet levels) to partition the feature space in way that maximize the discrimination ability for NN training ability to construct proper weight vectors that correctly classify the training set within some defined error rate. While, the trained classifier, in testing phase, assigns the unknown input pattern to one of the class encoding (person encoded ID) based on the extracted feature vector. Training and testing operations are performed using cross validation technique.

V. RESULTS

In this paper, the data set consists of different sounds recorded from hundred different persons (seventy males and thirty females) using a normal and common microphone (commercial microphone). Each Person is requested to say two different statements in order to save them in the data base and use it for training. In total, the data set consists of 200 samples (60 female and 140 male). Sentences that have been recorded are:

- My name is <Name>.
- I live in <City>.

To reduce the information lost of a speech signal, the parameter of the data acquisition should be selected according to the nature of the speech signal to be processed. The speech signal in this paper is sampled with 8 KHz, and quantized with 16-bit quantization level. These specifications use the minimum memory usage of speech signal, and thus, minimize the computational power that is required, without any distortion on the voice signal.

Table-1 shows the results of the 7 levels of wavelet decomposition with MLP recognizer. From the table, the columns represent different percentages of voice size (i.e. 50% means, the use of 50% of the coefficients of the wavelet transform). The rows represent different levels of wavelet decomposition. Where, the intersection of any row and column means the percent of accuracy.

TABLE I. VALIDATION RESULTS.

Data Percent / **Wavelet Level**	**100 %**	**70 %**	**50 %**	**40 %**	**30 %**	**20 %**
Level 1	65	65	65	61	55	30
Level 2	67	67	67	62	55	30
Level 3	98	98	98	94	84	70
Level 4	88	88	88	82	70	61
Level 5	92	92	92	85	74	60
Level 6	70	70	70	52	40	24
Level 7	53	53	53	38	25	12

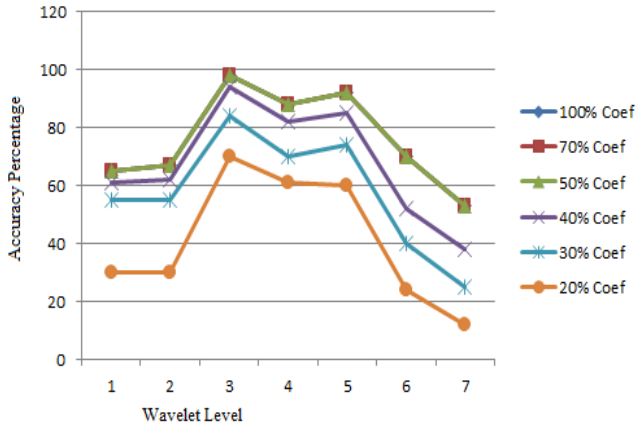

Figure 4. Validation Results.

Figure 4 shows the testing validation results using cross validation. The Y-axis represents the accuracy that gotten by validation. Where, X-axis represents the level of multi-level wavelet decomposition.

VI. CONCLUSION

This paper studies the effect of multi-level wavelet decomposition as feature extraction method in order to get the best condition of person's identification / verification depending on voice recognition.

The use of wavelet transformation in multi-level decomposition enables to represent the meaningful features of the human voice, in low size coefficients data and omitting the whole most of unwanted data in the human voice speech. From the other hand, determining the best level to work on future head researches is the job of this paper.

The system was been implemented and tested using cross validation on different 100 persons. From the results above, it's clear that; level three of the decomposition is the best level and gets the highest accuracy followed by level five. But level five comprise lower data size and thus, lower computational power and higher speed.

The use of whole coefficients of wavelet levels decomposition can be minimized by using 50% of it. That is clear from the results 50% of the coefficients locate the same voice information that is can be gotten helpfully from 100% of it. Thus, we can use 50% to minimize the computational power and increasing up the running speed.

Ones, the coefficients become less that 50% of its original decomposition, the human voice data starts to slow down and thus, the accuracy decreases significantly.

REFERENCES

[1] Russell Kay, "Biometric Authentication, Technical Report", CSO, the resource for security executives, 2005.

[2] A.L. Graps, "An Introduction to wavelets", IEEE Computational Science and Engineering, Vol 2. No. 2, pp. 50-61 1995

[3] George Tzanetrakis, Georg Essl, Perry Cook, "Audio Analysis Using the Discrete wavelet Transform" , Proceedings of WSEAS conference in Acoustics and music Theory Application , 2001.

[4] R.V Pawar, P.P. Kajave, and S.N. Mali, "Speaker Identification Using Neural network" , WASET, Vol.12, No.7, PP. 31-35, 2005.

[5] Chabane Djeraba, Hakim Saadane, "Automatic Discrimination in Audio Documents" , Nantes University, 2 rue dela Hossiniere, Bb 92208-44322 Nates Cedex3, France, 2000, pp.1-10.

[6] Evgeny Karpov. "Real-Time Speaker Identification", Master thesis. University of Joenssu, Department of Computer Science, 2003.

[7] Brain J. Love, Jennifer Vining, Xuening Sun, "Automatic Speaker Recognition Using Neural Networks", Technical report, the University of Texas at Austin, 2004.

[8] F Murtagh, J.L Strack. O Renaud, "On Neuro-Wavelet Modeling", School of Computer Science, Queens University Belfast, Northern Ireland, France, 2003.

[9] Roberto Gemello, Franco Mana, Dario Albesano. "Hybrid Hmm/Neural Network Based Speech Recognition In Loquendo" , ASR, 2006.