
***A PROPOSED SYSTEM TO DEFINE STUDENT IDENTITY THROUGH SOUND
RECOGNITION TECHNIQUES***

By

M. E. El-Alami

*Vice Dean for Post Graduate
Studies and Research, Head of
Computer Science Department
Mansoura University, Egypt*

M.F. El Atwi

*Computer Science Department
Mansoura University, Egypt*

M. A. Ezzat

*Computer Science Department
Mansoura University, Egypt*

Research Journal Specific Education

Faculty of Specific Education

Mansoura University

ISSUE NO. 36, OCTOBER. 2014

مجلة بحوث التربية النوعية – جامعة المنصورة

العدد السادس والثلاثون – أكتوبر ٢٠١٤

A PROPOSED SYSTEM TO DEFINE STUDENT IDENTITY THROUGH SOUND RECOGNITION TECHNIQUES

*M. E. El-Alami**

*M.F. El Atwi***

*M. A. Ezzat****

Abstract

This paper presents a framework for a proposed Sound recognition system to identify sound of human and detect who is the speaker. The presented system is an approach for identification sound through Wavelet and pattern recognition. Primary goals of the proposed system were to identify speaker sound by improve performance results and identify the sound. According to many researches and studies to the acoustic signature of many properties which make them the most discriminating between the dynamic characteristics of the human body in the identification and discrimination of sound, so everyone has a unique sound like one, everyone unique nervous system controls the audio device, everyone has their different sound waves from the rest of humanity. Firstly, the system receives sound from the person and converts it to digital signal. Secondly, it makes some pre-processing on signal to makes it clear by removing noise with the following functions: (I) Input Sound; (II) pre-processing; (III) signal feature extraction using wavelet analysis; (IV) classifying signal. And (V) detect the speaker; this showed in the diagram of the proposed system.

Keywords

Sound Recognition, Wavelet, Speaker Identification, Mel-Frequency Cepstral Coefficients (MFCC), feature extraction, recognition, signal recognition.

1. INTRODUCTION

Due to the significant changes taking place in the world community with the era of information and communication technology revolution, the programmers of educational institutions face significant challenges in their statutes and their regulations and methods, curricula, and educational

* Vice Dean for Post Graduate Studies and Research, Head of Computer Science Department Mansoura University, Egypt

** Computer Science Department Mansoura University, Egypt

*** Computer Science Department Mansoura University, Egypt

institutions in the Arab world need to revisit and seek permanent development to keep pace with these changes, and many States have taken on the need to review the current situation and adapted to be compatible with the age of information technology, the education system should benefit from the earnings of computer and communications technology. The main goal of this research is extract features of sound and describing an audio signal in terms of the sound events or sources that identify the sound acoustic. Sound recognition is an important search field for developing the main future aims such as robotics, security systems, content-based indexing of multimedia files, and human-machine interfaces. Most sound recognition research is aimed at improving one of these application domains, such as speech recognition [1] or music genre classification [2] [3] [4]. The methods in these application domains have proven successful in problems with a single, known type of sound, and even in recognizing isolated environmental sounds, such as footsteps or jangling keys [5]. However, these methods have some attributes that make them less suitable for sound recognition in real-world environments. The system that has to recognize sound events in a real-world environment cannot rely on the assumption that the input consists of a single, known, and undistorted signal type, as the methods used on speech and music often can. [6]

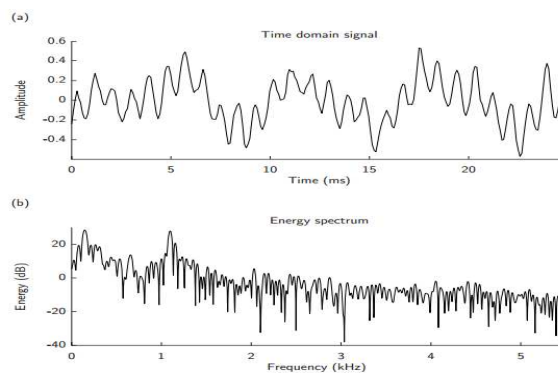


FIGURE 1: (A) THE AMPLITUDE SPECTRUM (B) THE AUDIO SIGNAL

A conventional speaker recognition system illustrated in Figure 2 is comprised of two stages: the first stage is called the enrolment or training process; the second stage is called the recognition or testing process. [6]

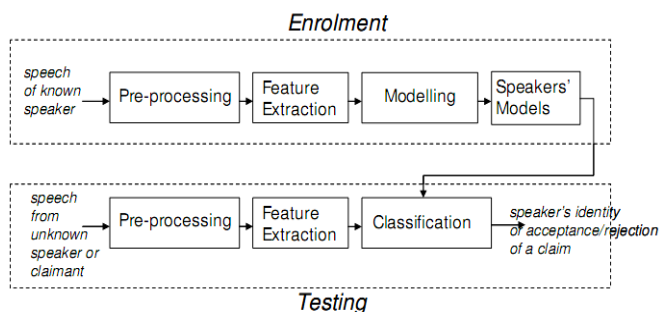


FIGURE 2: MAJOR COMPONENTS OF A CONVENTIONAL SPEAKER RECOGNITION SYSTEM [7]

2. RELATED WORK

This section shows some necessary literature of identifying student identity through sound recognition techniques.

L. Ma, T. Tan (2004) has presented an efficient algorithm for iris recognition, which is invariant to translation, scale and rotation. This method regards the texture of the iris as a kind of transient signals and uses the wavelet transform to process such signals. The local sharp variation points, good indicators of important image structures, extract from a set of intensity signals to form discriminating features. Experimental results have illustrated the encouraging performance of the current method in both accuracy and speed. In particular, a comparative study of existing methods for iris recognition has been conducted. Such performance evaluation and comparison not only verify the validity of our observation and understanding of the characteristics of the iris but also will provide help for further research.[8]

D. Peralta et al. (2014) have present a Fingerprint matching effective tool for human recognition due to the uniqueness, universality and invariability of fingerprints. The performing fingerprint identification over a large database can be an inefficient task due to the lack of scalability and high computing times of fingerprint matching algorithms. They proposed a distributed framework for fingerprint matching to tackle large databases in a reasonable time. It provides a general scheme for any kind of matcher, so that its precision is preserved and its time of response can be reduced. The

System has been tested they conduct an extensive study that involves both synthetic and captured fingerprint databases, which have different characteristics, analyzing the performance of three well-known minutiae-based matchers within the designed framework. With the available hardware resources, our distributed model is able to address up to 400000 fingerprints in approximately half a second. [9]

Atulya Velivelli et al. (2003) have proposed a general method to retrieve part of the sound using hidden Markov models and synthesis, the Portal provides the user query the voice clip through one or more of the parts of the volume, is the main feature of this portal based on hidden Markov models that assumed no boundaries defined by, and can retrieve the section of audio in a specific subject more accurately and effectively.[10]

M.Bahoura (2009) has proposed pattern recognition methods to classify respiratory sounds into normal and wheeze classes. he evaluate and compare the feature extraction techniques based on Fourier transform, linear predictive coding, wavelet transform and Mel-frequency Cepstral coefficients (MFCC) in combination with the classification methods based on vector quantization, Gaussian mixture models (GMM) and artificial neural networks, using receiver operating characteristic curves. He proposed the use of an optimized threshold to discriminate the wheezing class from the normal one. Also, post-processing filter is employed to considerably improve the classification accuracy. The Experimental results show that our approach based on MFCC coefficients combined to GMM is well adapted to classify respiratory sounds in normal and wheeze classes. McNemar's test demonstrated significant differences between the results obtained by the presented classifiers ($p < 0.05$).[7]

A. Drygajlo (2012) has discussed some important aspects of forensic speaker recognition, focusing on the necessary statistical-probabilistic framework for both quantifying and interpreting recorded voices as biometric evidence. Methodological guidelines for the calculation of the evidence and its strength under operating conditions of the casework were presented. He gave an example, an automatic method using the Gaussian

mixture models (GMMs) and the Bayesian interpretation (BI) framework were implemented for the forensic speaker recognition task. The BI method represents neither speaker verification, nor speaker identification. These two recognition techniques cannot be used for the task, since categorical, absolute and deterministic conclusions about the identity of source of evidential traces are logically untenable because of the inductive nature of the process of the inference of identity. The method, using a likelihood ratio to indicate the strength of the biometric evidence of the questioned recording (trace), measures how this recording scores for the suspected speaker model compared to relevant non-suspect speaker models. [11]

A. P. A. Broeders (2001) has detected that the development of state-of-the-art speaker recognition systems has shown considerable progress in the last decade, performance levels of these systems do not as yet seem to warrant large-scale introduction in anything other than relatively low-risk applications. Conditions typical of the forensic context, such as differences in recording equipment and transmission channels, the presence of background noise and of variation due to differences in communicative contexts continue to pose a major challenge. Consequently, the impact of automatic speaker recognition technology on the forensic scene has been relatively modest and forensic speaker identification practice remains heavily dominated by the use of a wide variety of largely subjective procedures. While recent developments in the interpretation of the evidential value of forensic evidence clearly favor methods that make it possible for results to be expressed in terms of a likelihood ratio, unlike automatic procedures, traditional methods in the field of speaker identification do not generally meet this requirement. However, conclusions in the form of a binary yes/no-decision or a qualified statement of the probability of the hypothesis rather than the evidence are increasingly criticized for being logically flawed. Against this background, the need to put alternative validation procedures in place is becoming more widely accepted.[12]

Sheeraz Memon (2010) the goal of the research was to investigate three major areas of improvements of the existing speaker recognition methodology. (1) To propose an improved modeling and classification

methodology for speaker recognition: This aim was achieved by the development of a new algorithm for the calculation of Gaussian Mixture Model parameters called Information Theoretic Expectation Maximization (ITEM). The results showed an improvement of the equal error rate over the classical EM approach. The EM-ITVQ also showed higher convergence rates compared to the EM. (2) To determine the usefulness of features derived from nonlinear models of speech production for speaker recognition. This aim was achieved by comparing the classical features based on linear models of speech production with recently introduced features based on the nonlinear model. New fusions of features carrying complimentary speaker-dependent information are proposed. The speaker verification experiments presented demonstrated significant improvement of performance when the conventional MFCC features were replaced by a fusion of the MFCCs with complimentary linear features such as the inverse MFCCs (IMFCCs), or nonlinear features such as the TMFCCs and TEO-PWP-Auto-Env. The higher overall performance of the nonlinear features when compared to the linear features was observed. (3) To determine the effects of a clinical environment containing clinically depressed speakers on speaker recognition rates, and to investigate if the features based on nonlinear models of speech production have the potential to counteract the inverse effects of the clinically depressed environment.[13]

This paper motivates to Submission of a proposed technology improves the performance of the security systems for the protection of data and information through the design of the proposed system to identify persons thought sound recognition system. The structure of the proposed system is shown in the following section.

3. METHODOLOGY

The proposed system module consists of the following stages: (1) Sound acquisition, (2) Sound pre-processing; (3) Feature Extraction analysis; and (4) Classification. Fig. 1 shows the main block diagram of the proposed system.

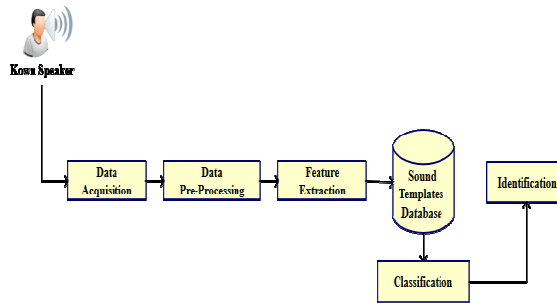


FIGURE 3:THE PROPOSED SYSTEM COMPONENT BLOCK DIAGRAM

3.1.1 Sound Acquisition

The first step of sound recognition is data acquisition. In the real world, the analog data from the physical world are acquired through a transducer, and further digitized to discrete format suitable for computer processing.[14] We knew that to input sound you must have the following:

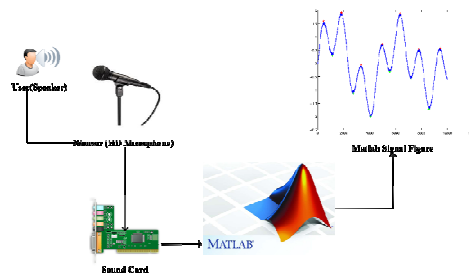


FIGURE 4: ACQUISITION CYCLE FROM SPEAKER SOUND TO MATLAB SIGNAL ANALYSIS

In Acquisition stage we record 10 patterns for each 10 Speakers in the proposed system by using "Sound Record" tab as shown in the following:

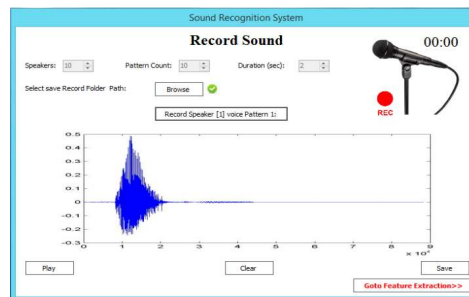


FIGURE 5: RECORDING SOUND IN THE PROPOSED SYSTEM GUI

3.1.2 Sound Pre-Processing:

Audio Pre-processing (APP) is the task of partitioning and classifying an audio stream [15]. A major function of the data pre-processing part is to modify the measured data obtained from the data acquisition part so that those data can be more suitable for the further processing in feature extraction and classification. There are many modifications in the data pre-processing part will show as the following:

❖ *Truncation:*

The default sampling frequency of wavread command is 44100 Hz. [16] when an audio clip is recorded, say for duration of 2 secs, the number of samples generated would be around 90000 which are too much to handle. Hence we can truncate the signal by selecting a particular threshold value. We can mark the start of the signal where the signal goes above the value while traversing the time axis in the positive direction. In the same we can have the end of the signal by repeating the above algorithm in the negative direction.

❖ *A Pre-Emphasis*

The pre-emphasis filter is typically a simple first order high pass filter. The purpose of pre-emphasis is to even the spectral energy envelope by amplifying the importance of high-frequency components and removing the DC component in the signal. The z-transform of the filter is shown in this Eq.: [17].

$$H(z) = 1 - \alpha z^{-1}, 0.9 \leq \alpha \leq 1.0$$

In the time-domain, the relationship between the output s'_n and the input s_n of the pre-emphasis block. For fixed-point implementations a value of $\alpha = 15/16 = 0.9375$ is often used (Rabiner & Juang, 1993). [18]

$$s'_n = s(n) - \alpha s(n-1)$$

The pre-emphasized signal was divided into short frame segment using Hamming window. [17]

The process of splitting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within the

range of 20 to 40 msec. The voice signal is divided into frames of N samples.[16],[19] The continuous speech signal is divided into frames of N samples, with adjacent frames being separated by M samples with the value M less than that of N. The first frame consists of the first N samples. The second frame begins from M samples after the first frame, and overlaps it by N - M samples and so on. This process continues until all the speech is accounted for using one or more frames. We have chosen the values of M and N to be N = 256 and M = 128 respectively. Figure 3.3. Below gives the frame output of the truncated signal.[16]

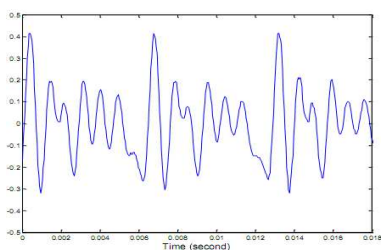


FIGURE 6: FRAME OUTPUT SIGNAL

❖ *Windowing:*

To reduce the distortion in a spectrum which occurs at the first and last part of the frame windowing concept is used. Here, hamming window is related in this equation

$$W(i)=0.54- 0.46 \cos(2\pi i/ T-1), 0\leq i\leq T-1 [16]$$

Where T is the Number of Frames

❖ *Fast Fourier Transform*

The fast Fourier transform (FFT) is used extensively in many scientific fields, but the canonic FFT is afflicted by two pitfalls, the spectral leakage due to time limitation, and the picket fence effect (PFE) due to discrete frequencies in the calculated spectrum. Actually, these two pitfalls were found shortly after the FFT was presented by Cooley et al.[20]. Because of the extensive application of the FFT, these pitfalls have been causing concern for some time. One of many developed techniques for alleviating the pitfalls is called windowing. Numerous references report that the two pitfalls can be alleviated by delicately choosing windows [21],

though not eradicated completely. This is because using windows the discrete spectral lines are viewed as non-parametrically, that is, it does not consider the relationship between neighbor lines of the FFT spectrum [22].

STFT is represented in the discrete domain given by this Eq. :

$$X[m, k] = \sum_{n=0}^{N-1} x[n] W[n - m] e^{-jkn/N} \quad [23]$$

where $W[n]$ is a short-time windowing function of size L , centered at time, location m , and N is the number of discrete frequencies ($N \geq L$). Usually, N is chosen to be a power of -2 of using an efficient, fast Fourier transform (FFT). Since the Fourier transform is a complex function, the power spectrum density (PSD) is used and is given by this Eq. :

$$P_s[m, k] = \frac{1}{N} |X[m, k]|^2 \quad [23]$$

❖ **Cepstrum:**

The Cepstrum of a signal is the Fourier transform of the logarithm of its power spectrum. Let $X(\omega)$ denotes the spectrum of the voiced speech signal, $P(\omega)$ denotes the spectrum of the pitch impulses and $H(\omega)$ denotes the spectrum of the vocal tract which includes the effects of glottal wave form. The relation between the magnitudes of these three spectra can be expressed simply as follows: [24]

$$|X(\omega)| = |P(\omega)| * |H(\omega)|$$

Taking the logarithm of this equation gives:

$$\log\{|X(\omega)|\} = \log\{|P(\omega)|\} + \log\{|H(\omega)|\}$$

Thus, in the logarithm of $|X(\omega)|$, the contributions due to $P(\omega)$ and $H(\omega)$ are added. The contribution from $H(\omega)$, which is essentially determined by the properties of vocal tract itself, tends to vary slowly with frequency, while the contribution from $P(\omega)$ tends to vary more rapidly and periodically with frequency. These two components should be separated by

means of a linear filtering operation. Taking the Fourier transform of it, Cepstrum can be obtained. [24]

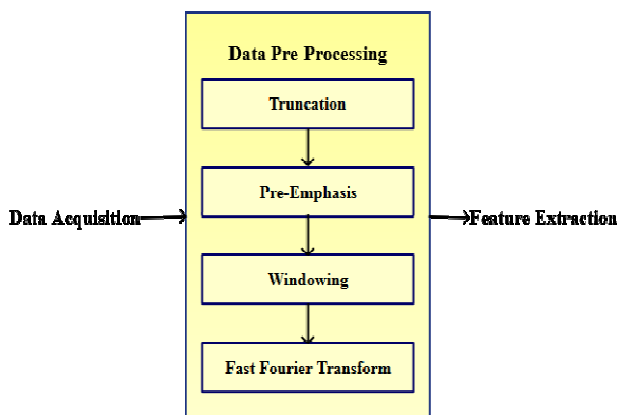


FIGURE 7:PRE-PROCESSING STAGE

3.1.3 Feature extraction

Feature extraction is the heart of a signal recognition system. In Signal recognition, features are utilized to identify one class of signal from another. The signal space is usually of high dimensionality. The objective of the feature extraction is to characterize the input sound and further, to reduce the dimensionality of the measurement space to a space suitable for the application of signal classification techniques. Feature extraction can be viewed as a mapping, which maps a pattern space into a feature space, and the dimensionality of the feature space has to be smaller than pattern space.[14]

In Brief we can say that feature extraction is the automated process of locating and encoding distinctive characteristics from a biometric sample in order to generate a template. [25] And it's mainly the process that transforms originally high-dimensional vectors into lower dimensional vectors. It is a mapping $f: \mathbb{R}^N \rightarrow \mathbb{R}^D$, where $D < N$. Feature extraction can be considered as a data reduction process that attempts to capture the essential characteristics of the analyzed signal with a small data rate. This operation is described in figure 8, where the analyzed signal is segmented in frames using short and overlapped window functions.[7]

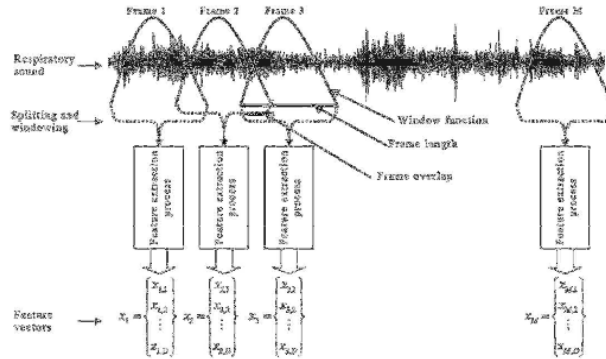


FIGURE 8: BLOCK DIAGRAM OF A TYPICAL FEATURE EXTRACTION PROCESS. THE ANALYZED SIGNAL IS SEGMENTED IN FRAMES USING SHORT AND OVERLAPPED WINDOW FUNCTIONS. EACH FRAME IS THEN CHARACTERIZED BY A REDUCED SIZE FEATURE VECTOR [7]

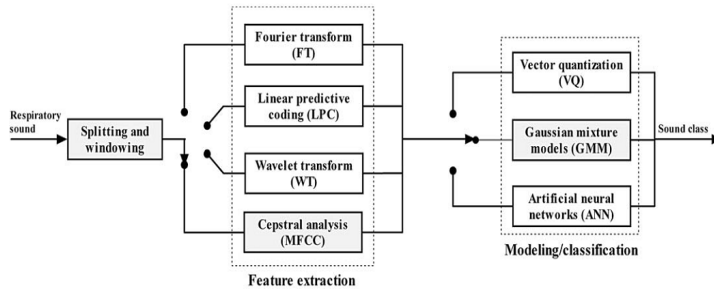


FIGURE 9: SOUND CLASSIFIERS OBTAINED THROUGH VARIOUS COMBINATIONS OF FEATURE EXTRACTION AND CLASSIFICATION METHODS [7]

❖ **Fourier Transform (FT):**

The short-time Fourier transform (STFT) of a discrete-time n is defined as

$$S[m, k] = \sum_n s[n]w[n - m]e^{-j2\pi nk/N} \quad [7]$$

Where $w[n]$ is a short-time windowing function of size L , centered at time location m and N is the number of discrete frequencies. Since the Fourier transform is a complex function, we use the power spectrum density (PSD) given by

$$P_s[m, k] = \frac{1}{N} |S[m, k]|^2$$

At the sampling frequency f_s , each windowed signal (frame) is represented by N -points power spectrum covering the frequency range $[-f_s/2, f_s/2]$. The power spectrum cannot be used directly as feature vector because it contains a large size data ($N/2$ components).

To classify asthmatic breath sounds, *Rietveld et al.* [6] compute the power spectra from shorter intervals of approximately 3s, representing a full breathing cycle. The frequency range 100-1300 Hz is divided into 26 bands of approximately 46 Hz each. The feature vector of 26 components is obtained by averaging its power spectrum in each band. In this approach, noted R-FT method, the feature vector is defined by

$$\mathbf{x} = [\bar{P}_1, \bar{P}_2, \dots, \bar{P}_{26}]^T$$

where P_k is the average power spectrum in the k th band and T is the transpose operation.

Fig. 4 gives an example of power spectrum of breath sounds obtained from both healthy and asthmatic subjects. There are significant differences between spectra in the upper frequency.

❖ *Linear predictive coding (LPC):*

Linear predictive coding (LPC), also called auto-regressive (AR) modeling, is widely used in signal processing and specially in speech processing applications. The linear prediction model assumes that each sample can be approximated by a linear combination of a few past samples:

$$\hat{s}[n] = \sum_{k=1}^p a_k s[n-k], \quad n = 1, 2, \dots, N$$

Where $S[n]$ is the prediction of the signal $s[n]$, and the vector $\mathbf{a} = [a_1, \dots, a_p]$ is the coefficients vector of a predictor of order p . The prediction error $e[n]$ for n the sample $x[n]$ is given by the difference between the actual sample and its predicted value:

$$e[n] = s[n] - \sum_{k=1}^p a_k s[n-k]$$

The predictor coefficients a_k are solved by minimizing the mean-square value of the estimation error.

Sankur et al. [9] used a 6th order AR-model in combination with a k-nearest neighbor (fc-NN) classifier in order to distinguish pathological respiratory sounds from healthy respiratory sounds. In this approach, named S-AR method, the feature vector is constructed from the LPC coefficients a_k and mean-square prediction error eq:

$$\mathbf{x} = [a_1, a_2, \dots, a_p, \varepsilon_p]^T$$

❖ **Wavelet Transform (WT):**

Wavelet analysis is originally introduced in order to improve seismic signal analysis by switching from short-term Fourier analysis of new better algorithms to detect and analyze abrupt changes in signals Daubechies. Wavelet may be seen as a complement to the classical Fourier decomposition method. Suppose, a certain class of functions is given and we want to find ‘simple functions’ f_0, f_1, f_2, \dots . Such that each:

$$f(x) = \sum_{n=-\infty}^{\infty} a_n f_n(x) \quad [26]$$

for some coefficients a_n .

Wavelet mean ‘small wave’. So wavelet analysis is about analyzing signal with small duration finite energy functions. They transform the signal under investigation into another representation which converts the signal in a more useful form. This transformation of the signal is called wavelet transform. Wavelet transform have advantages over traditional Fourier transform for representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic and non-stationary signals. Unlike Fourier transform, we have a different type of wavelets that are used in different fields. Choice of a particular wavelet depends on the type of application in hand. If the process is done in a smooth and continuous fashion, then transform is called continuous wavelet transform. If the scale and position are changed in discrete steps, the translation is called discrete wavelet transform. Note that in case of Fourier transform, spectrum is one-dimensional array of values whereas in wavelet transform, we get a two dimensional array of values.

Also note that the spectrum depends on the type of wavelet used for analysis. Mathematically, we denote a wavelet as:[27]

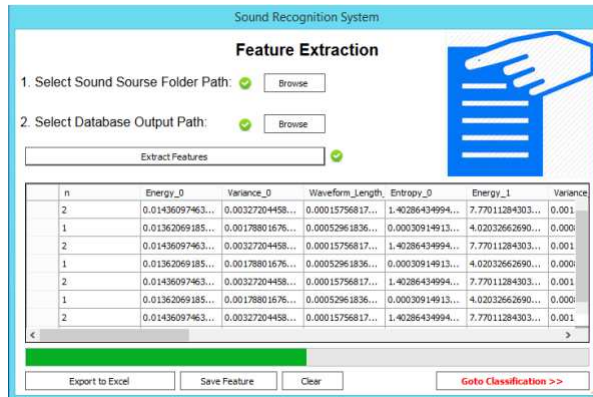
$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in R, a \neq 0,$$

Where b is location parameter and it is scaling parameter. For the function to be wavelet, it should be time limited. For a given scaling parameter a, we translate the wavelet by varying the parameter b.

The wavelet energy is the sum of square of detailed wavelet transform coefficients [28]. The energy of wavelet coefficient is varying over different scales depending on the input signals. The wavelet energy of coefficients c(t) can be defined as follows:

$$E(s(t)) = \sum_{j=1}^N a_j c_j^2 \quad [29]$$

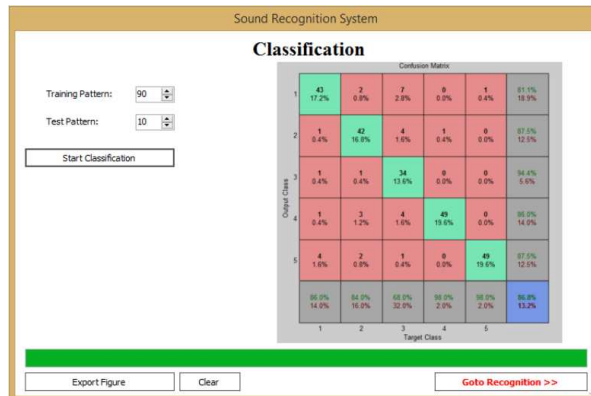
The main advantage of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time–frequency resolution in all frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets lack of the requirement of stationary [30]. The proposed System Tab that used to extract features:



3.1.4 Classification:

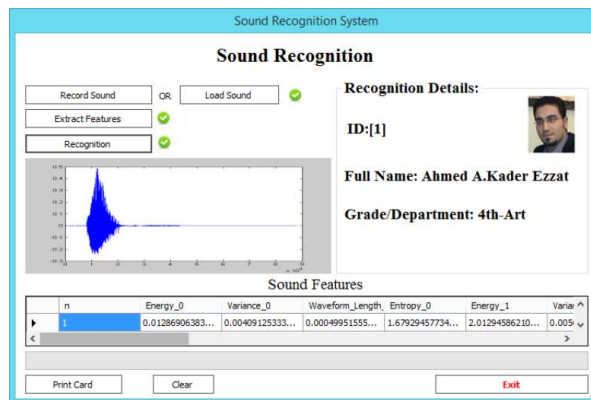
In this stage we classify the sound database to test and training best spread factor can lie between 0.06 and 0.10 to get maximum classification

accuracy using ANN and Pattern Matching Algorithm classifier. In this stage, the proposed pattern recognition system is trained and tested using the sound database and their corresponding between sound files. By using classification tab in GUI we can classify the database to test and trainer shown:



3.1.5 Identification:

In this stage we start by record voice using "Record" Button after that we click "Extract Feature" Button to extract the query sound feature and draw sound and show query sound figure tabulated in the bottom of the GUI.



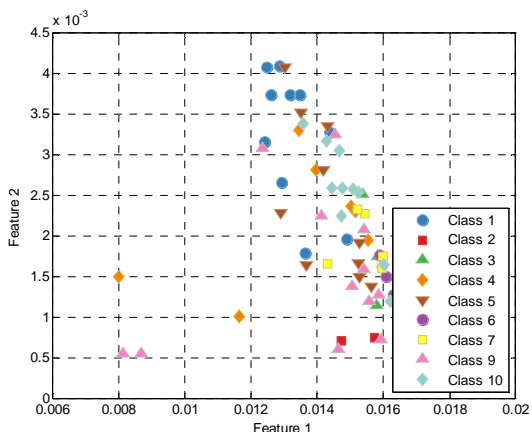
4. Experimental Work:

This research was tested on 100 pattern sample that Acquisition by high definition Microphone and save it in sound database by using proposed system GUI and use the wavelet transform to extract feature for 100 pattern.

Then we classify this database and make 10 test speaker patten sample and 90 training pattern samples. After that we make query to detect and identify speaker for unknown speaker the proposed system find the match from the database.

❖ **Dataset Visualization:**

The sound feature dataset contains 20 features, the wavelet levels and energy of subband, from 10 different persons. To create classification dataset is created that has the same observations but has binary targets that indicate if an observation is of the certain class as shown:



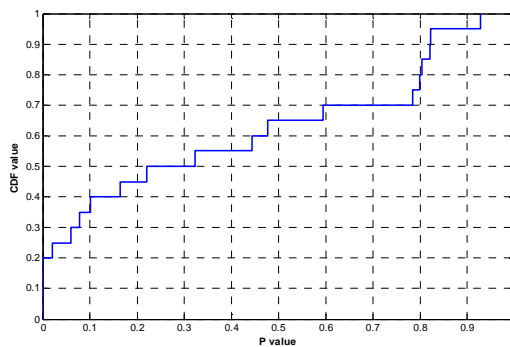
Finally, the explore method of the prtDataSets enables interactive visualization of different dimensions of the data sets using GUI controls. You can start the explore GUI using the command below, and you'll be greeted by the GUI Figure below.

❖ **Feature selection:**

Selecting features for classifying high-dimensional data reducing the number of features (dimensionality) is important in statistical learning. For many data sets with a large number of features and a limited number of observations, such as bioinformatics data, usually many features are not useful for producing a desired learning result and the limited observations may lead the learning algorithm to over fit to the noise. Reducing features can also save storage and computation time and increase comprehensibility.

There are two main approaches to reducing features: feature selection and feature transformation. Feature selection algorithms select a subset of features from the original feature set; feature transformation methods transform data from the original high-dimensional feature space to a new space with reduced dimensionality.

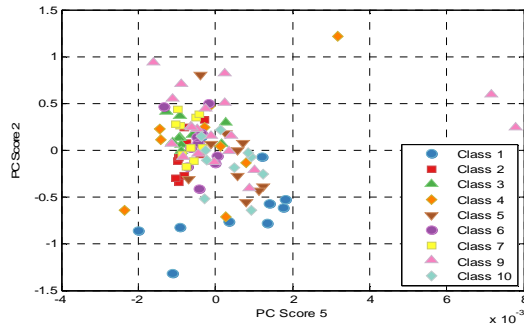
This demo focuses on feature selection techniques. More specifically, this demo shows how to use the functions in the Statistics Toolbox(TM) to perform sequential feature selection, which is one of the most popular feature selection algorithms. It also shows how to use holdout and cross-validation to evaluate the performance of the selected features.



❖ *Data Processing:*

Now that we've explored some of the techniques available to visualize data, let's apply some transformations to the data. To do this, we need to know a little bit about prtActions. Most everything you can do to data in the PRT - algorithms/classifiers/regressors in the PRT are implemented as prtActions.

For now, all that's important to know is that prtActions have two important methods - "train" and "run". The train method accepts one prtDataSet and outputs an object of the same class as the prtAction being used. The "run" method, in contrast, accepts a prtDataSet input and outputs another prtDataSet with data changed to reflect the action undertaken.

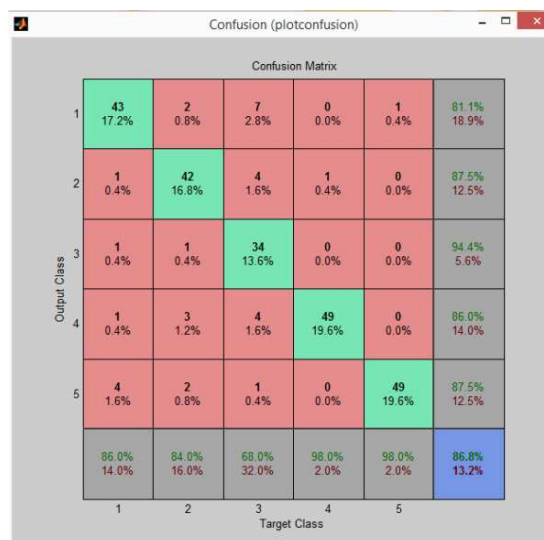


❖ **Building a Classifier:**

Let's continue processing our new data set in PCA space. Let's say we'd like to generate a classifier that can tell the difference between the sounds in the data set. Since the sound data set has multiple classes, we need to use a classifier that can handle multiple hypothesis data. KNN classification algorithms are a decent choice in this case. Let's build and visualize a KNN classifier on the 3-PC projected sound data.

❖ **Evaluating the Classifier:**

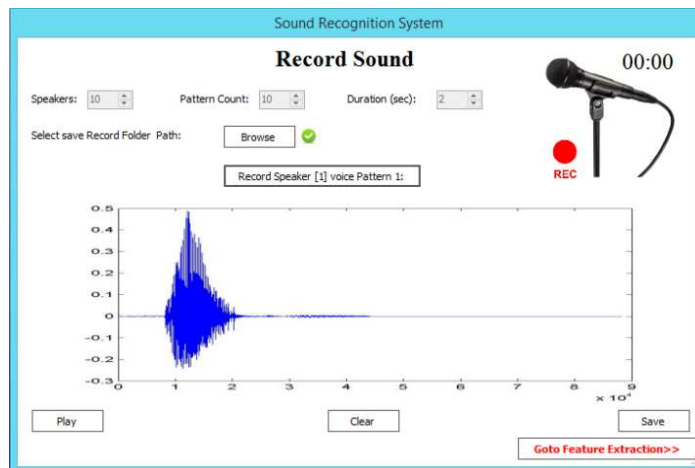
A display of the confusion matrix appears that shows various types of errors that occurred for the final trained network. The next figure shows the results.



5. Conclusion:

The proposed program which in current research has been tested on 10 speakers and the results are high-accuracy up to 86.8%. The sound recognition approach is a fabulous field to search in. We must also highlight the importance of the use of sound in the identification of individuals. In this paper, a computer-based sound recognition system has been developed define students in the college. The proposed system can acquire, save, analyze, and recognize the speaker. The proposed system includes two main modules: (1) Record Sound and pre-processing; (2) signal feature extraction using wavelet analysis; (3) classifying signal to test & train and (4) Identification to detect who is speaker. Firstly, the system receives the sound from the unknown person and converts it to the signal. Secondly, it receives the sound database to extract features. Then classify it to detect the person. The Results shown that the proposed system is effective in identify student identity.

The Proposed System GUI



Sound Recognition System

Feature Extraction

1. Select Sound Source Folder Path:

2. Select Database Output Path:

Extract Features

n	Energy_0	Variance_0	Waveform_Length	Entropy_0	Energy_1	Variance
2	0.01436097463...	0.00327204458...	0.00015756817...	1.40286434994...	7.77011284303...	0.001
1	0.01362069185...	0.00178801676...	0.00052961836...	0.00030914913...	4.02032662690...	0.000
2	0.01436097463...	0.00327204458...	0.00015756817...	1.40286434994...	7.77011284303...	0.001
1	0.01362069185...	0.00178801676...	0.00052961836...	0.00030914913...	4.02032662690...	0.000
2	0.01436097463...	0.00327204458...	0.00015756817...	1.40286434994...	7.77011284303...	0.001
1	0.01362069185...	0.00178801676...	0.00052961836...	0.00030914913...	4.02032662690...	0.000
2	0.01436097463...	0.00327204458...	0.00015756817...	1.40286434994...	7.77011284303...	0.001

Sound Recognition System

Classification

Training Pattern:

Test Pattern:

Confusion Matrix

1	43 17.2%	2 0.8%	7 2.8%	0 0.0%	1 0.4%	81.1% 18.9%
2	1 0.4%	42 16.8%	4 1.6%	1 0.4%	0 0.0%	87.5% 12.5%
3	1 0.4%	1 0.4%	34 13.6%	0 0.0%	0 0.0%	94.4% 5.6%
4	1 0.4%	3 1.2%	4 1.6%	49 19.6%	0 0.0%	95.0% 4.9%
5	4 1.6%	2 0.8%	1 0.4%	0 0.0%	49 19.6%	87.5% 12.5%
	95.0% 14.0%	94.0% 16.0%	68.0% 32.0%	98.0% 2.0%	98.0% 2.0%	86.8% 13.2%
	1	2	3	4	5	

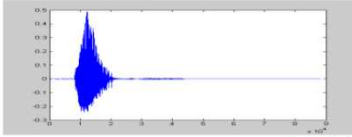
Sound Recognition System

Sound Recognition


Record Sound Load Sound

Extract Features

Recognition



Recognition Details:

ID:[1] 

Full Name: Ahmed A.Kader Ezzat

Grade/Department: 4th-Art

REFERENCES

- [1] J. O'SHAUGHNESSY and e. al., "A randomized study of lapatinib alone or in combination with trastuzumab in heavily pretreated HER2+ metastatic breast cancer progressing on trastuzumab therapy," *J Clin Oncol*, vol. 26, no. 15S, p. 1015, 2008.
- [2] B. L. STURM, "Classification accuracy is not enough," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 371-406, 2013.
- [3] J. NEAL, "Music Recommender Systems and Genre Bias," 2012.
- [4] J.-J. AUCOUTURIER and F. PACHET, "Representing musical genre: A state of the art," vol. 32, no. 1, pp. 83-93, 2003.
- [5] M. COWLING and R. SITTE, "Comparison of techniques for environmental sound recognition," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2895-2907, 2003.
- [6] M. Niessen, "Context-Based Sound Event Recognition," Ph.D, Rijksuniversiteit Groningen, 2010.
- [7] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Computers in Biology and Medicine*, vol. 39, no. 9, p. 824–843, September 2009.
- [8] T. T. L. Ma, "Efficient iris recognition by characterizing key local variations," *IEEE Trans. Image Process*, vol. 13, p. 739–750, 2004.
- [9] I. T. R. S.-R. F. H. a. J. B. D. Peralta, "Fast Fingerprint Identification for Large Databases," *Pattern Recognition*, vol. 47, no. 2, p. 588–602, February 2014.
- [10] A. Velivelli, C. XiangZhai and T. S. Huang, "Audio Segment Retrieval Using a Synthesized HMM," Beckman Institute for Advanced Science and Technology, 2003.

- [11] A. Drygajlo, "Automatic Speaker Recognition for Forensic Case Assessment and Interpretation," School of Criminal Justice, Federal Institute of Technology Lausanne, University of Lausanne, 2012..
- [12] A.P.A.Broeders, "Forensic Speech and Audio Analysis Forensic Linguistics," 13th INTERPOL Forensic Science Symposium conference, 16-19 October 2001.
- [13] S. MEMON, "Automatic Speaker Recognition: Modelling, Feature Extraction and Effects of Clinical Environment," PhD Thesis, 2010.
- [14] G.-C. PAN, "A Tutorial of Wavelet for Pattern Recognition".
- [15] H. D. DOS SANTOS MEINEDO, "Audio Pre-processing and Speech Recognition for Broadcast News," PhD Thesis, 2008.
- [16] D. Sabitha, R. Parameshwaran, K. Hariharan, V. Vaithyanathan and S. Backialakshmi, "Text Independent Speaker Recognition," International Journal of Applied Engineering Research, vol. 8, no. 20, 2013.
- [17] O. Chia Ai, M. Hariharan, S. Yaacob and L. Sin Chee, "Classification of speech dysfluencies with MFCC and LPCC features," Expert Systems with Applications, vol. 39, no. 2, pp. 2157-2165, 2012.
- [18] L. R. RABINER and B.-H. JUANG, Fundamentals of speech recognition, Englewood Cliffs: PTR Prentice Hall, 1993.
- [19] S. B. Magre and R. R. Deshmukh, "Design and Development of Automatic Speech Recognition of Isolated Marathi Words for Agricultural Purpose," IOSR Journal of Computer Engineering, vol. 16, no. 3, pp. 79-85, 2014.
- [20] J. Cooley and J. Tukey, "An algorithm for the machine computation of complex Fourier series," Mathematics of Computation, vol. 19, p. 297-301, 1965.
- [21] H. Gaberson, "A comprehensive windows tutorial," Sound and Vibration, vol. 40, p. 14-23, 2006.

- [22] Y. F. Lia and K. F. Chenb, "Eliminating the picket fence effect of the fast Fourier transform," *Computer Physics Communications*, vol. 178, no. 7, p. 486–491, April 2008.
- [23] M. HARIHARAN, R. SINDHU and S. YAACOB, "Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network," *Computer methods and programs in biomedicine*, vol. 108, no. 2, pp. 559-569, 2012.
- [24] I. M. EL-HENAWY and e. al., "Recognition of phonetic Arabic figures via wavelet based Mel Frequency Cepstrum using HMMs," *HBRC Journal*, vol. 10, no. 1, pp. 49-54, 2014.
- [25] M. A. Dawson, "The Use of Technology to Combat Identity Theft," *Study Conducted Pursuant to Section 157 of the Fair and Accurate Credit Transactions Act of 2003*, 2005.
- [26] M. Sifuzzaman, M. R. Islam and M. Z. Ali, "Application of Wavelet Transform and its Advantages Compared to Fourier Transform," *Journal of Physical Sciences*, vol. 13, pp. 121-134, 2009.
- [27] M. KUMAR and S. PANDIT, "Wavelet transform and wavelet based numerical methods: an introduction," *International journal of non linear science*, vol. 13, no. 3, pp. 325-345, 2012.
- [28] F. MÖRCHEN, "Time series feature extraction for data mining using DWT," *Germany*, 2003.
- [29] S. EKICI, S. YILDIRIM and M. POYRAZ, "EKICI, Sami; YILDIRIM, Selcuk; POYRAZ, Mustafa. Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2937-2944, 2008.
- [30] D. AVCI, "An expert system for speaker identification using adaptive wavelet sure entropy," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6295-6300, 2009.