Research Journal of Applied Sciences, Engineering and Technology 13(9): 717-721, 2016
DOI:10.19026/rjaset.13.3345
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: March 25, 2016 Accepted: July 14, 2016 Published: November 05, 2016

Research Article

Speaker Identification Using Evolutionary Algorithm

Dr. Jane J. Stephan
University of Information Technology and Communication, Baghdad, Iraq

Abstract: The aim of this study was identifying the speaker using evolutionary algorithm such as Genetic Algorithm (GA). This study provides an efficient approach for speaker identification using Discrete Wavelet Transform (DWT) and the Energy in feature extraction stage and Genetic Algorithm used in classification stage to recognize the sound of speaker. Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. The files that have been chosen here named WAV and recording sounds with different speakers. In practical side was used Delphi language (version 7) to tests the course. The results showed that the recognition good was 87%.

Keywords: Energy, genetic algorithm, speaker identification, wavelet transform

INTRODUCTION

Speaker recognition is the process of instantly recognizing who is speaking on the basis of individual information part of speech waves. This technique accomplish it possible to use the speakers' say/tell to authenticate their identity and ruling access to services such as say/tell dialing, banking by telephone, telephone shopping, database access services, information services, say/tell mail, security ruling for confidential information areas and ancient access to computers. In this study, the speaker recognition concepts and techniques are described (Ong *et al*., 1996). Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker (Rostem, 1999).

Speaker recognition methods can also be divided into text-independent and text dependent methods. In a text-independent system, speaker models take characteristics of somebody's speech which show up irrespective of what one is saying. In a text-dependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking one or more specific phrases enjoy passwords, card numbers, PIN codes, etc. (Furui, 2001).

All technologies of speaker recognition, identification and verification, text independent and text-dependent, each have their own benefits and disadvantages and May requires different treatments and techniques. The choice of which science to use is app-specific.

At the highest level, all speaker recognition systems have two main modules: feature extraction and feature races. Feature extraction is the process that extracts a little amount of data from the say/tell signal that can after be used to describe each speaker. Feature races involves the true plan to identify the infamous speaker by comparing extracted features from his/her say/tell input with the ones from a set of famous speakers (Kinnunen *et al*., 2000). The objective of this study is providing method for identification the speaker by recognize the sound of speaker without regardless the word spoken using one of the evolutionary algorithms.

LITERATURE REVIEW

Tolba (2011) in this study report an approach that depends on Continuous Hidden Markov Models (CHMMs) to identify Arabic speakers automatically from their voices. The Mel-Frequency Cepstral Coefficients (MFCCs) were selected to describe the speech signal. The general Gaussian density distribution HMM is developed for the CHMM system. Ten Arabic speakers were used to evaluate our proposed CHMM-based engine. The identification rate was found to be 100% during text dependent experiments. However, for the text-independent experiments, the identification rate was found to be 80%.

Sarkar and Saha (2010) in this study, the computational complexity and identification time mainly depend on the number of speakers, the number of frame vectors, their dimensionality and the model

order of the classifier. Due to the slow movement of the voice producing parts, the adjacent frame vectors do not very much in information content. In this study, we present the design of a speaker identification system with a distance metric based frame selection technique. The aim is not only to provide the architecture of a speaker identification system but also to reduce the redundant frames at the pre-processing stage to lower the identification time and computation burden which are vital for real time implementation (Sarkar and Saha, 2010).

Kumari *et al*. (2012) in this study provides an efficient approach for text-independent speaker identification using a fused Mel feature sets and Gaussian Mixture Modeling (GMM). The individual Gaussian components of a GMM are shown to represent some speaker specific spectral shapes that are effective for modeling speaker identity. Two different set of features which are complement to each, other, Mel Frequency Cepstral Coefficient (MFCC) and Inverted Mel Frequency Cepstral Coefficient (MFCC) features are obtained for each speaker and are trained using Expectation Maximization algorithm. Two GMM models; one for MFCC feature sets, other one for IMFCC feature sets are created. During testing phase, the likelihood of unknown speaker's features with each of the GMM models is determined. By using a weighted sum of these likelihood values, a fused score is created for each speaker and speaker with maximum score is the identified speaker. The performance of this fusion GMM system is evaluated using a part of the TIMIT database consisting of 120 speakers and a maximum identification efficiency of 93.88% is achieved.

Saeed and Nammous (2007) in this study discusses a Speech-and-Speaker (SAS) identification system based on spoken Arabic digit recognition. The speech signals of the Arabic digits from zero to ten are processed graphically (the signal is treated as an object image for further processing). The identifying and classifying methods are performed with Burg's estimation model and the algorithm of Töeplitz matrix minimal eigenvalues as the main tools for signal-image description and feature extraction. At the stage of classification, both conventional and neural-network-based methods are used. The success rate of the speaker-identifying system obtained in the presented experiments for individually uttered words is excellent and has reached about 98.8% in some cases. The miss rate of about 1.2% was almost only because of false acceptance (13 miss cases in 1100 tested voices). These results have promisingly led to the design of a security system for SAS identification. The average overall success rate was then 97.45% in recognizing one uttered word and identifying its speaker and 92.5% in recognizing a three-digit password (three individual words) (Saeed and Nammous, 2007).

GENETIC ALGORITHM

Genetic algorithm based on natural genetics; therefore they portion identical names. The genetic algorithms is a stochastic search technique (stochastic search use probability to aid administer their search) inspired by the mechanics of natural selection and natural genetics (Goldberg, 1989).

The basic thought unhurried the genetic: An algorithm is to hold a population of strings or chromosome, which are encoding of a possibilities solution to the quandary being investigated. Each chromosome is tested using a fitness function to know the excellent solution of the quandary. The new population is created by selecting chromosome from the old population. The new population is re-evaluated and the processes abide until the solution is found (Mitchell, 1996).

The strings of artificial genetic advancement are analogous to chromosome in biological advancements. The chromosomes are quiet of features, or detectors that are called genes. This may hold on some number of values, called alleles. Features may be located on alternative spaces on the string, the quandary of genes, it is locus, is identified separately from the gene's function. Thus, we can talk a particular gene, for example, an animal's eye color gene, its locus, space 10 and its allele value, blue eyes. The total package of strings (chromosome) is called a structure. These structures decoded to create a particular parameter set (Mitchell, 1996).

In natural advancements, one or more chromosome combined to create the total genetic prescription for the construction and operation of some organism. The total genetic package (structure) is called the genotype. The organism put together by the interaction of the total genetic package with its environment is called the phenotype (Maouche and Benmohamed, 2009).

WAVELET ANALYSIS

Wavelet transform is a description of a signal in terms of set of basic functions, which is obtained by dilation and translation of a basic wavelet. Since wavelets are brief time oscillatory functions having finite encourage length (limited duration both in time and frequency), they're localized in both time (special) and frequency domains. The joint spatial-frequency attachment obtained by wavelet transform assemble it a wonderful candidate for the extraction of details as celebrated as approximations of pictures (Fan *et al*., 2009).

THE PROPOSED SYSTEM

In this section, an expose for the recognition continuity has been demonstrated as follows.

The dataset contents: The course has been applied on eight English words; these words were recorded by microphone with text-dependent and independent speaker (5 speakers) two men and three women. Each speaker repeats each word 6 times, three of these words are storing as reference file in database and the other three words are used for testing. The format of these files are wave format The total number of words in the database becomes 240 utterances,120 utterances are used for storing as reference and 120 utterances are used as testing in the proposed algorithm. Each word has different length. These words are (file-open-save-close-copy-cut-paste-shutdown). This word was recorded at sampling good of 22 KHz coded in 8 bits, one contact The digitization of the signal was made by the professional software "sound makes version 9.0" (Saeed and Nammous, 2007).

Architecture of the proposed system: The advancement is an independent speaker (multi speaker), includes three stages, the preprocessing stage, the feature extraction stage and the classification stage, Fig. 1 shows the process of proposed advancement.

Preprocessing stage: This stage represents the signal processing share (i.e. converting the signal to its parametric representation this stage include Sampling to be usable by a computer, the signal must be transition from analog sound to the digital sound (Holmes and Holmes, 2001). And framing it speech signal is blocked into adjusts of m samples. Since we arrangement with speech signal, which is non-stationary signal (vary with time), the framing process is vital to arrangement with adjust not with orderly signal. Back this stage the speech.

Signal has many adjusts and the number of adjusts depends on the number of samples for each word. The number of samples in each adjust is 256 samples.

Feature extraction stage: The speech signal consisting of infinite information, we must extract the most critical ones. An advice comparison treatment on this kind of signal is impossible because there is too distinguished information. So the Discrete Wavelet Transform (DWT) using Haar function coefficients are applied on all the arranges for all words used in this proposed work, the filter bank used for the extraction of features is (three and four) level, The DWT of the extracted arrange (256 samples) aspiration befall in four sub bands called (d1, d2, d3, a3). Since most of the information is concentrated in the low frequency components only the 32 samples befall from the low be a live filters (a3) aspiration is considered as the features of the input signal. The DWT using (Haar) scaling function coefficients are applied on all these arranges for all words used in this proposed work. Then the energy for each signal to invent befall ant features also named feature vectors that are classified by GA in classification stage.

The energy of the speech signal provides a convenient picture that reflects these amplitude variations. Eq. (1) shows the computation of energy in each arrange. This stage is used to prick the number of data characterizing and shows a diminutive number of parameters or coefficients (Rabiner and Schafer, 1978):

$$E_n = \sum_{n=-\infty}^{\infty} [x(n)]^2 \quad (1)$$

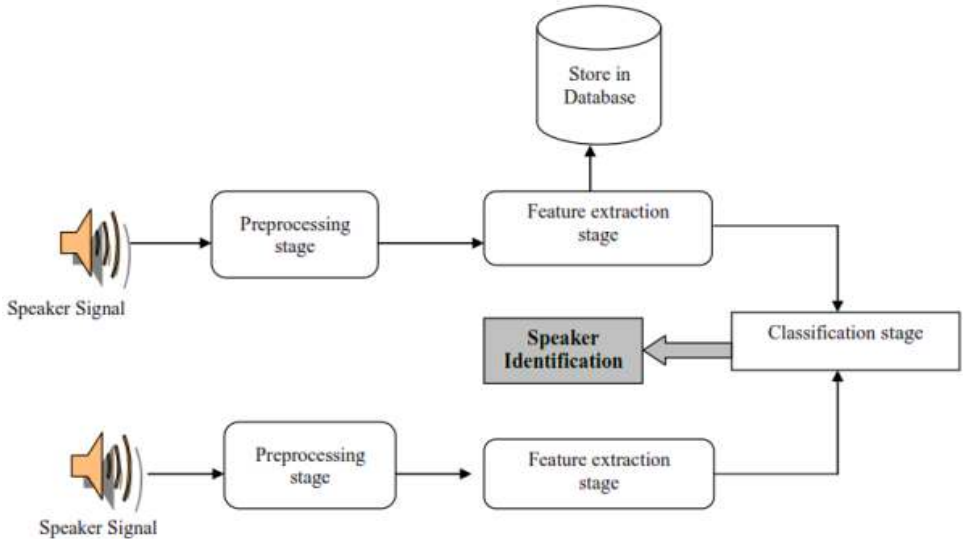

Fig. 1: The proposed method

Table 1: Energy of each sound file

Word (Open)	Energy				
Test file before process	25738	75767	75029	53890	38898
In binary	110010010001010	10010011111110111	10010010100010101	1101001010000010	1001011111110010
Reference file	84846	51165	19293	17523	16351
In binary	10100101101101110	1100011111011101	100101101011101	100010001110011	11 111111110111
After crossover	110010010001010	10010011111110111	10010010100010101	1101001010000011	1001011111110011

where,
n : Number of samples
X (n) : The value of signal

Classification stage: The classification stage is made by a genetic algorithm by comparing the test and reference files to identify the speaker.

Initialization of the population: The reference files are the population managed by Genetic Algorithms, the number of the files in the population is 160 files (individuals). The choice of the initial population is random for each word and for different speakers to be recognized. Binary encoding is used to encode the energy of feature vector for speech signal.

Evaluate the population: To evaluate the population, the fitness of each individual or chromosome are be calculated, which is the differences between the speaker signal to be recognized and each speaker signal in the database, whenever the distance value is less or approach to zero then the speaker is recognized. Its formula is derived from the Euclidean distance using gauge Square Error as shown in Eq. (2) (Armstrong and Collopy, 1992):

$$Distance\ (A, B)=\frac{1}{N}\sum_{i=1}^{n}\sqrt{(A-B)}^{2} \quad (2)$$

A : The vector test file
B: The vector reference file
N: Number of block

Selecting: After evaluating individuals of the population, the elitist selection scheme admiration be used; this scheme allows the Genetic Algorithm to a number of best individuals for the back generation. These individuals may be adrift if they are not selected to reproduce.

Crossover: Single site crossover was used in Magnitude features block by replace the value of some blocks between the dispute and reference that are selected randomly, the probability of the crossover is 0.7. Table 1 presents the energy of some speech segment for both dispute and reference file and crossover operation between them and the file back of crossover.

Table 2: Identification rate

Gender	Identification rate (%)
Male	79
Male	83
Female	95
Female	91
Female	87

Mutation: A mutation operation in this bond is simple change because keeping the feature extraction is very indispensable in this bond, so if the speaker did not recognized, the mutation operation appetite be done by replacing the individual. The probability of mutation is 0.001.

The stop criterion: The stop criterion of proposed continuity is either the speaker is recognized or all sub-populations acquire been covered or the number of generation is finished.

IMPLEMENTATION AND RESULTS

The proposed procedure was used recorded sound using sound bring about program, with different speaker (two men and three women) and eight different words, these words selected because using in conduct computer. The format of file sound is WAV with sampling suited 22 KHZ. Table 2 shows the suited of identification of speaker for all words mentioned in database contents.

CONCLUSION

In this study, we suggest an arrangement for Speaker Identification using genetic algorithm that was used for recognition. The recognition of the speaker was text-dependent reveal 8 English words each word recording 6 times. Our furtherance used Discrete Wavelet Transforms (DWT) to increases the efficiency of the statistical calculations due to the decrease in speech area calculations and energy in feature extraction stage. Also we use a genetic algorithm in classification stage with binary encoding and single crossover. The recognition suitable when is 87%.The furtherance was implemented by Delphi programming language (version 7).

REFERENCES

Armstrong, J.S. and F. Collopy, 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. Int. J. Forecasting, 8(1992): 69-80.

Fan, C.N., H.B. Wang and F.Y. Zhang, 2009. Improved wavelet-based illumination normalization algorithm for face recognition. Proceeding of the 1st International Conference on Information Science and Engineering (ICISE, 2009), pp: 583-586.

Furui, S., 2001. Digital Speech Processing, Synthesis and Recognition. Marcel Dekker, New York.

Goldberg, D., 1989. Genetic Algorithm in Search, Optimization and Machine Learning. Addison Wesley, Longman, Boston, MA.

Holmes, J. and W. Holmes, 2001. Speech Synthesis and Recognition. 2nd Edn., Taylor and Francis e-library London and New York.

Kinnunen, T., T. Kilpelinen and P. Frinti, 2000. Comparison of clustering algorithms in speaker identification. Proceeding of the IASTED International Conference on Signal Processing and Communications (SPC, 2000). Marbella, Spain, pp: 222-227.

Kumari, R.S.S., S. Selva Nidhyananthan and G. Anand, 2012. Fused Mel feature sets based text-independent speaker identification using Gaussian mixture model. Proc. Eng., 30: 319-326.

Maouche, F. and M. Benmohamed, 2009. Automatic Recognition of Arabic Word by Genetic Algorithm and MFFC Modeling. Faculty of Informatics, Mentouri University, Constantine, Algeria.

Mitchell, M., 1996. An Introduction to Genetic Algorithms. The MIT Press, Cambridge, MA.

Ong, S., S. Sridharan, C.H. Yang and M.P. Moody, 1996. Comparison of four distance measures for long time text-independent speaker identification. Proceeding of the 4th International Symposium on Signal Processing and Its Applications (ISSPA, 1996), pp: 369-372.

Rabiner, L. and R.W. Schafer, 1978. Digital Processing of Speech Signals. Prentice-Hall, Englewood Cliffs, New Jersey.

Rostem, H., 1999. Speaker identification using multi-layer neural network. M.Sc. Thesis, University of Babylon, Iraq.

Saeed, K. and M.K. Nammous, 2007. A speech-and-speaker identification system: Feature extraction, description and classification of speech-signal image. IEEE T. Ind. Electron., 54(2): 887-897.

Sarkar, G. and G. Saha, 2010. Real time implementation of speaker identification system with frame picking algorithm. Proc. Comput. Sci., 2: 173-180.

Tolba, H., 2011. A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach. Alexandria Eng. J., 50(1): 43-47.