Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.11(3) 2019 , pp Comp 21–30

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Advantages and Disadvantages of Automatic Speaker Recognition Systems

Rawia Ab. Mohammed [a] , **Akbas E. Ali** [b] , **Nidaa F. Hassan** [c]

[a] *Department of Computer Science , University of Technology, Baghdad, Iraq.Email: rawyia_8080@yahoo.com*

[b] *Department of Computer Science, University of Technology , Baghdad, Iraq.Email: akbas_ali2006@yahoo.com*

[c] *Department of Computer Science , University of Technology, Baghdad , Iraq, Email: 110020@uotechnology.edu.iq*

ARTICLE INFO

Article history:
Received: 05 / 05 /2019
Rrevised form: 26 /05 /2019
Accepted : 16 / 06/2019
Available online: 26 /06/2019

Keywords:

Automatic Speaker Recognition, Voice Biometrics, Deep Learning, Adaptive MFCC.

ABSTRACT

Automatic speaker recognition systems use the machines to recognize an individual via a spoken sentence. Those systems recognize a specific individual or confirm an individual's claimed identity. The most common type of voice biometrics is the Speaker Recognition. Its task focused on validation of a person's claimed identity, using features that have been obtained via their voices. Throughout the last decades a wide range of new advances in the speaker recognition area have been accomplished, but there are still many problems that need solving or require enhanced solutions. In this paper, a brief overview of speech processing is given firstly, then some feature extraction and classifier techniques are described, also a comparative and analysis of some previous research are studied in depth, all this work leads to determine the best methods for speaker recognition. Adaptive MFCC and Deep Learning methods are determined to be more efficient and accurate than other methods in speaker recognition, thus these methods are recommend to be more suitable for practical applications.

1. Introduction

Processing of speech has appeared as a significant application field of digital signal processing. A great number of areas for researching in processing of speech keep emerging like speech identification, speaker identification, speech coding, and speech synthesis. Speaker identification aims to recognize automatically the speaking person based on particular information that is included in the waves of the speech. This approach utilized the voice of the speaker for verifying her/his identity and provides control access to services like voice dialing, information services, Data-base access

Corresponding author Rawia Ab. Mohammed

Email addresses: Rawyia_8080@yahoo.com

Communicated by Dr. Mustafa Jawad Radif

services, voicemail, remote access to PCs, security control for private information fields, and a number of other areas in which security is a considerable subject of interest [1]. A system of speaker identification generally includes two main phases, which are: training and testing [2]. The initial stage of Training phase includes extraction of parameters from the speech signal for obtaining a representation which is appropriate for statistical modeling, therefore models are commonly utilized in the majority of speaker identification systems. The next stage includes obtaining a statistical model from the parameters. The test phase work as follows; the system entries are the claimed identity, and the samples of speech which are uttered by an unidentified person. As known, the aim of a speaker identification problem is verifying whether or not the speech samples match the claimed identity. Parameters of speech undergo extraction from the speech signal with the use of precisely the same module which is utilized for training. After that, the speaker model which corresponds to the claimed identity will be obtained from the group of statistical models that have been computed throughout training. Then, the obtained speech parameters and the statistical model calculate scores and recognizes who the speaker actually is [3], [4]. The presented research is organized in the following manner; section II described state of art with advantages and disadvantages of some important techniques, and section III presents a description of two types of speaker recognition. Finally, conclusion for this paper is presented in section IV.

2. Related Works

Speaker identification approaches are divided to text-dependent or–independent. In the case of the former, the same enrollment and verifying test are used, while in the latter, the text throughout enrolling and test are differs. Text-independent systems are usually utilized for speaker recognition due to the fact that they need quite little cooperation from the individual [5].

In the present paper some important speech recognition methods are reviewed and presented as follows:

In [6], authors used Quantum Neural Networks technique (QNN) with an Improved Particle Swarm Optimization (IPSO), IPSO-QNN is used to improve the recognition performance of speech recognition, and QNN characteristics are:

Growing memory capacity and speed; smaller sizes of networks and simpler topologies for those networks; fast study and high-speed information processing abilities; more sufficient stability and reliability; in addition to the ability to eliminate catastrophe of amnesia [7]. *This technique provides better accuracy rate and consumed time, which make it best for faster quantum neural computation and PSO's global improvement.*

In [8], authors described the development of speech recognition system utilizing Weight Mel Frequency Cepstral Coefficients (WFCC) and Dynamic Time Warping (DTW) techniques. The system is split to 2 stages, which are training and recognition, in training phase, speaker recorded the words at the speed slower than normal speech rate, and save them for the original template, and WFCC feature is extracted. In the recognition phase, words is recorded using normal speech rate, extracted the WFCC feature at the same time, and then match them using the DTW algorithm. *The results refers that WFCC had a better rate of recognition in isolated words From MFCC.*

In [9], MFCC and inverted MFCC are used, in which performance is enhance for speaker identification , since it's containing high frequency area complementary data, also the Gaussian shaped filter (GF) is utilized to calculate MFCC and inverted MFCC instead of conventional triangle shaped bins. The basic concept is introducing more correlation between the outputs of the sub band. The efficiency of each of MFCC and inverted MFCC is improved with Gaussian Filter over the conventional triangular filter (TF) based implementation, separately and combined. The approach of Vector Quantization (VQ) feature matching has been utilized, because of its high precision and simplicity. The suggested research had accomplished 98.57% of accuracy.

In [10], two techniques for Speaker Recognition are used. The first technique is Artificial Neural Networks (ANN) which is used to train the features vectors that obtained after preprocessing speech signal, and the second technique is "Particle Swarm Optimization" (PSO) for improving these features. The PSO finds the best areas of the complicated search space via the interaction of persons in the population. The obtained results achieved a good accuracy.

In [11], three technique are presented: Wavelet Octave Coefficients of Residues (WOCOR), Gauss Mixture Model (GMM) and Naïve Bayes Classifier. WOCOR for capturing the spectrotemporal source excitation properties which are embedded in the Linear Predictive (LP) residual signal. The information of "vocal Source" (WOCOR) includes pitch frequency and phase in the residual signal. WOCOR is produced via the application pitch synchronous wavelet transformation to the residual signal. Pitch Synchronous wavelet transformation is utilized properties of the excitation signal for capturing the spectrotemporal.

In [12], a developed system of speech identification is utilized various approaches like MFCC, VQ and Hidden Markov Model (HMM). MFCC is utilized for the properties extraction of input speech signal based on a certain word that is spoken by a specific speaker, HMM utilized Quantized feature vectors for the identification of the word via the evaluation of the maximum log possibility values for the uttered word.

In [13], a speech identification approach is described based on VQ and Improved Particle Swarm Optimization (IPSO). This approach contained the following stages: de-noising, feature mining, VQ, and IPSO based Hidden Markov Model approach (IP-HMM). Initially, the speech signals are de-noised with the use of median filter, then, properties are obtained from the de-noised signal. Next, the obtained properties are granted to genetic algorithm based code-book generation in vector quantizing. The initial populations are produced via choosing arbitrary code vectors from the training dataset for the code-books for the process of the genetic algorithm. IP-HMM is helpful for recognizing. The presented speech identification approach reaches an accuracy rate of about 97.14%.

In [14], mathematical algorithm is present for Real Time Speaker Identification System by utilizing MFCC, VQ and LBG (Linde, Buzo and Gray) algorithms. The phase of speaker identification benefits from MFCC and VQ, while the phase of speaker recognition utilizes LBG algorithm. This technique provides 91% of precision in normal conditions.

In [15], voice recognition is suggested to develop speech identification system. This work had made use of Adaptive MFCC and Deep Learning techniques. Adaptive MFCC is used for the reduction of data loss, and imposing the weighted value to the data region. *Thereby, it avoids loss of data that results improvement of the rate of identification. Moreover, the Deep Learning presents the possibility of using the voice identification with no Database (DB).*

In [16], authors described the input speech signal via the recognition of the content involving: pre-processing, feature extraction and Multiple Support Vector Machine (SVM). For optimizing those features, various optimizing methods have been used like Adaptive PSO (APSO) approach. Those optimal features are presented as the input to multi SVM. From the results the optimization algorithm (APSO) obtained 97.8% precision.

In [17], a real-time speaker gender identification system has been designed and it was capable of identifying the gender of a speaker via his/her voice automatically. This system used FFT for preprocessing, mathematical calculations of voice frequency and voice fundamental frequency for features extracted. Machine learning classification technique (Naive Bayesian Classifier) is used for classification, this technique provides good performance about 92.75% for gender recognition.

Feature extraction, classification techniques, accuracy of classification and advantages and disadvantages of the above Speaker Recognition System are described in Table 1.

Ref. No.	Year	Feature Extraction Techniques	Classification Techniques	Database	Accuracy	Advantages	Disadvantages
[6]	2009	MFCC and Wavelet-Spectral compression	IPSO- QNN	The samples have been logged by ten people, every one of which repeats the voice ten times.	MFCC= 85% Wavelet= 84.5%	- These techniques combine parallel, fast QNN Computation with an ability of IPSO' global optimizing, QNN' computation quantity of training process is small. - The most significant benefit of the IPSO-QNN is its fast recognition. - IPSO considers precocious convergence phenomenon, when it is utilized for training QNN, the model predicts more accurately via fully exploiting the capability of global optimization of PSO. - MFCC captures main characteristics of phones in speech with low complexity.	- These techniques do not give the impression of parallel computing of quantum. - In background noise MFCC doesn't perform precise results and its efficiency could be influenced by the number of filters
[8]	2010	WMEL-Weight Frequency Cepstrum Coefficients (MFCC)	Dynamic Time Warping (DTW)	The samples are recorded by 10 male and 10 female voices sound by the recording software of PC Windows's system.	If number of words to be recognized is 3,the accuracy for system= 96.4%	- Combining these techniques led to increase speed of recognition for the isolated word, reduced storing space for the reference template and higher predictive accuracy. - DTW algorithm is easy and common hardware is required, so it is a more mature voice recognition algorithm. - DTW algorithm is to find an optimal time warping function which minimizes the total amount of the cumulative distortion. - WFCC have good recognition performance and noise immunity.	- This type of method of recognition module does not take under consideration the inherent speech variability. - Choosing the appropriate reference template is a difficult task. -which is why, the majority of modern systems of ASR replace this method (DTW) by a stochastic method like HMM.
[9]	2011	Inverted MEL-Frequency Cepstrum Coefficients (MFCC) XOR Gaussian shaped filter (GF)	Vector Quantization (VQ)	The samples are recorded by Microsoft sound recorder. There were 70 speakers of twenty utterance of each.	98.57% very short test voic sample 2 seconds.	- These techniques provided higher amount of correlation between sub band outputs. - Improve the efficiency of both MFCC and inverted MFCC by GF. - Using VQ with MFCC and Inverted are capable of detecting the precise speaker from considerably shorter samples of speech, which is why, it is well applicable in real-time applications. - The VQ is high accurate, simple, used less memory, less preparing of test data.	The VQ that it does not take into account the temporal evolution of signals because all the vectors are mixed up.

[10]	2011	Fast Fourier Transform (FFT)	PSO- ANN	Two speaker databases are used, training and testing. In every speaker data-base there are 11 speakers.		- The Results have a good precision, despite the relative randomness of the society of speakers. - Use of PSO method resulted in: Simple implementation, easy parallelization for concurrent processing, a rather small number of algorithm parameters and highly sufficient global search. - PSO can be a sufficient tool of optimization for non-linear continuous optimizing issues and combinatorial problems of optimization.	- The search direction is not obvious and it tends to have fast and early convergence in mid optimum points. -This technique cannot accommodate with the problems of non-coordinate system.
[11]	2012	Wavelet Octave Coefficients Of Residues (WOCOR)	GMM and Naive Bayesian	TIMIT data-base includes 630 Speakers.	GMM= 93.02% Naive Bayesian= 86.33%	- The Bayes classification represents an approach of supervised learning and also a statistical classification approach. - Bayes classification can provide a good viewpoint for the understanding and evaluation of several learning algorithms. It obtains explicit possibilities for hypothesis and it has robustness against noise in the input data.	Performance of the system decreases in the case of using GMM, and when training and testing environments are different.
[12]	2013	MEL-Frequency Cepstrum Coefficients (MFCC) ,and Distance Minimum algorithm	VQ and HMM	The samples are recorded by 4 people, every speaker is speak the same word ten times.	Speaker=S S1= 92% S2= 94% S3= 98% S4= 96%	-VQ is a highly sufficient spectral information representation in speech signals by vector mapping (feature extraction of input signal) from large vector space to a finite number of areas in the space which is referred to as clusters. - combining MFCC and Distance Minimum algorithm produced the optimal Performance and precise results. - HMM is capable of identifying the most widely utilized isolated word.	The VQ that it does not take into account the temporal evolution of signals because all the vectors are mixed up.
[13]	2014	Mathematical calculations of voice frequency.	VQ and (IP-HMM)	Two databases are used: First, It is trained with the use of 35 speech signals by 10 speakers. Second, It is trained with the use of 165 speech signals by 13speakers.	IP-HMM= 97.14%	- Combining between IPSO and HMM gives high accuracy compare with NN,HMM and PSO - IP-HMM is helpful in speech identification that is achieved via efficiently finding optimum or near optimum solutions in wide search spaces of code vector from code-book. - IPSO achieved global optimization ability.	- The search direction is not obvious and tendency to a fast and premature convergence in mid optimum points. -The VQ that it does not take into account the temporal evolution of signals because all the vectors are mixed up.
[14]	2015	MEL-Frequency Cepstrum Coefficients (MFCC)	VQ and LBG algorithm	The samples are recorded by 120 speakers with TIDIGIT database.	91%	- The VQ has greatly helpful to reduce the amount of computation and storage. - The VQ is high accurate, simple and less preparing of test data.	- Background noise MFCC does not give accurate results.

[15]	2016	Adaptive MFCC	Deep Learning	The 6 words like Hello, Turn on, Off, Up , Down , and bye are recoded and Utilized as data.	Hello= 98% Turn On=96% Turn Off=97.2% Up=97.6% Down=98% Good bye=97.4	- Deep Learning use small memory space, and improves the rate of identification. - Adaptive MFCC eliminated noise in every band with no damage to the audio via eliminating the small size noise via advanced smoothing in.	- Advanced adaptive filter is required for removing white noise especially for home appliances. - Deep Learning require a large amount of data and powerful computing resources.
[16]	2016	MEL-Frequency Cepstrum Coefficients (MFCC)	APSO and SVM	A group of words used as data.	APSO= 97.8%	-The advantage of APSO: They simplify the computation and present good results of recognition along with reduction in time utilization.	-This approach cannot work out the problems of non-coordinate system. - In background noise MFCC does not give accurate results.
[17]	2017	Mathematical calculations of voice frequency	Naïve Bayes	Speech was recorded into Mat lab directly	Naïve Bayes= 92.75%	Speaker recognition system is reliable.	

TABLE 1. Techniques for Speaker Recognition System

3. Types of Speaker Recognition

Speaker Identification Tasks or Speaker Recognition is a general term referring to tasks that discriminate between individuals according to the properties of their voices. Within this general description, there are 2 distinct tasks which were extensively researched. Those are termed as speaker recognition and verification. In some cases, the term "voice" or "talker" is replaced by "speaker," other cases, the term "authentication" is replaced by "verification." Speaker identification and voice authentication indicate the same task and based on those tasks that could construct two types of systems, which are [18]:

A. Speaker Identification System (SIS)

In this system task, it is focused on determine the speaker from a group of known speakers. It determines the person that has provided a certain sound according to the information which is contained in sound waves. As mentioned, the unknown audio is obtained from a certain group of known voices, this is why the task is known as the closed set identification. This process is a 1: N match in which the voice is compared with N templates Error which might happen in an SIS is the false identification of speaker. Speaker recognition System is illustrated in Figure 1

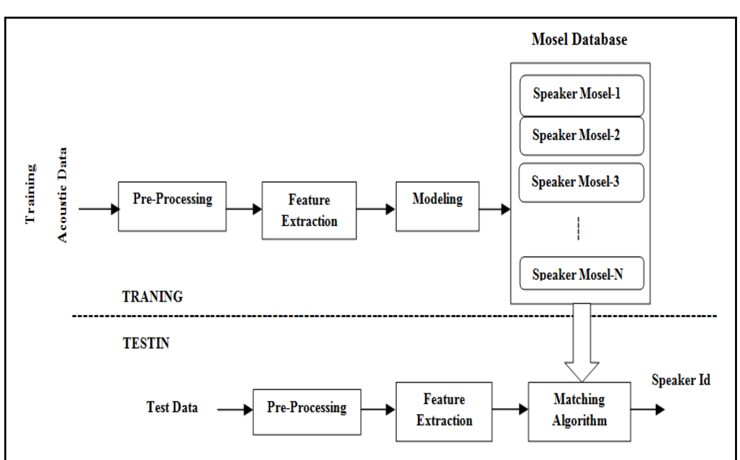

Fig 1. Speaker Identification System [5]

B. Speaker Verification System (SVS)

SVS is composed of five stages, they are: digital speech data acquisition, extracting features, pattern matching, decision making, as accept or reject, and enrolling for the generation of the models of speaker reference. A diagram of this process is depicted in Fig. 2. The stage of feature extraction maps every speech interval to a multi-dimensional feature space. This feature vector set is afterwards compared with speaker models via pattern matching, which produces a match score of every one of the vectors or set of vectors. The match score performs a measure of similarity of the calculated input feature vectors against models of the claimed speaker or their patterns of feature vectors. Lastly, a decision will be made for accepting or rejecting of the claimed speaker.

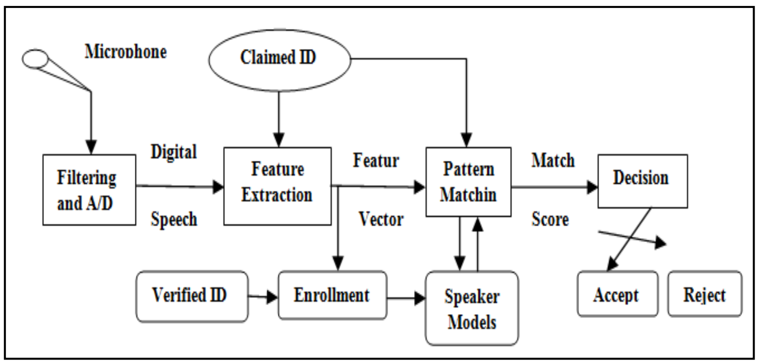

Fig 2. Speaker-verification system [19]

The following section is described the main phase of Speaker Recognition System:

A. Preprocessing

The recorded voice could include undesired background noises, unvoiced sounds, and there might also be a device mismatching, environmental mismatch between training and testing audio data that eventually degrades the efficiency of the SIS. The procedure of removing this noise, segmenting sounds to voiced and unvoiced sounds for enhancing speech/voice is referred to as preprocessing. The Preprocessing of voice data sample involves the following techniques [17]:

1- Pre-emphasis approach.
2- A/D conversion approach.
3- Time-to-Frequency domain conversion (FFT).
4- Frame blocking and hamming window.
5- End-point detection which includes the methods of noise and silence signal elimination.

B. Feature Extraction

Extracting speech features is a highly significant step for the transformation of speech signals to a group of feature vector coefficients containing only the information required to identify a specific sound. Due to the fact that each speech has different distinct properties that are enclosed in uttered words, those featured might be pulled out from many different methods for feature extraction and might be used for the task of speech identification. But pulled out feature must satisfy specific measures at the same time as dealing with the speech signals, for example: obtained features of speech must be computed easily, the obtained characteristics must be steady with time, and features must be robust against noise and environment [20]. Characteristics that are utilized for speaker identification may be classified by 3 main attributes, which are:

1- Discrete vs. continuous values
2- Temporal span
3- Information level

The information in speech signals occurs at a number of various rates and time spans. Which is why, features that are utilized for capturing this information occur as well with various rates and time spans.

The speech measurements value which is utilized in the features could be continuous or discrete. Features that consist of samples of speech frequency spectrum are an example of continuous valued measurements. Features counting the several occurrences of events in speech, like the counts of word usage, are example of discrete values measures. The third attribute represents the features of information level. Speech implies numerous information levels from semantic meaning through the words that are uttered to the physical vocal apparatus of the speaker, and through the acoustic sound of the speech [21].

Different feature extraction techniques are used for *SVS*, such as:

1- Mel-Frequency Cepstrum Coefficients (MFCC)
2- Perceptual Linear Prediction (PLP)
3- Linear Prediction Cepstral Coefficients (LPCC)
4- Discrete Wavelet Transform (DWT)
5- Linear Discriminant Analysis (LDA)
6- Linear Predictive Coding (LPC)
7- Principal Component analysis (PCA)
8- Relative Spectral (RASTA-PLP)

C. Modeling

Modeling is the process of creating template/speaker model from the features vectors which extracted from voice data sample. Modeling are categorized into two models, they are:

1) Template Model

The simplest model of this type has only one template ***x*** and it is the speech segment model. The match score between the template ***x*** for the claimed speaker and an input feature vector ***y*** from an unknown user.

2) Stochastic Model

It has been used recently, it provides a higher level of flexibility and produces a better score for matching. In this model. Pattern matching process is performed via calculating the probability of a feature vector in a certain speaker model.

D. Classification

Classification techniques for speaker identification in the past few years have been focused on statistical methods. The structure and selection of a classifier is dependent on the utilized techniques and features. The following is a list of subset classifiers which were successfully utilized in systems of automatic speaker identification [21]:

1- Support Vector Machine (SVM).
2- Sequence Kernels for Speaker Recognition.
3- Artificial Neural Networks.
4- Normalization Techniques.
5- Hidden Markov Modeling (HMM).
6- Gaussian Mixture Modeling (GMM).

4. Discussion of Results

In this paper, different feature extraction and classifier methods of speaker recognition through a study, comparison and analysis of previous researches are presented. Table (1) demonstrate the advantages and disadvantages of each method to recommend best method to be used in building and development an efficient speaker identification system. It's noticed that the best way to extract feature is the Adaptive MEL-Frequency Cepstrum Coefficients (MFCC), where noise is eliminated in every band with no damage to the audio, eliminating small size noise is done via advanced smoothing. In classification methods, the best recommend method is Deep

Learning, where it's used small memory space, and improves the rate of identification. As noted in [15], when these two methods were used together high accuracy was obtained in the rate of 98%.

CONCLUSION

The goal of automatic speaker identification is to acquire the voice of speakers and to create voice model for each speaker and finally to compares these models with an utterance of the speaker to prove his/her identity. There are a lot of techniques which have developed recently to improve the performance and accuracy of this technology, this paper reviewed types of speaker recognition and analyzed different types of feature extraction and classifier methods of speaker recognition used for identifying speakers. Advantages and disadvantages are also determined from reviewing the previous researches, it can be concluded that the best results were obtained when choosing adaptive MFCC for features extraction and Deep Learning for speech classification.

REFERENCES

[1] S. Gupta ; "Application of MFCC in Text Independent Speaker Recognition," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 5, May 2016 ISSN: 2277 128X .

[2] J. P. Campbell, "Speaker recognition: A tutorial," Proc. IEEE, vol. 85, no. 9, pp. 1437–1462, Sep. 1997.

[3] D. A. Reynolds; "Speaker identification and verification using Gaussian mixture speaker models," Speech Common., vol. 17, no. 1–2, pp. 91–108, 1995.

[4] Nisha.V.S and M.Jayasheela; "Survey on Feature Extraction and Matching Techniques for Speaker Recognition Systems," International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 2, Issue 3, March 2013.

[5] Z. Saquib, N. Salam, R. P Nair, N. Pandey and A. Joshi, "A Survey on Automatic Speaker Recognition Systems," *CDAC-Mumbai,Gulmohar Cross Road NO.9,Juhu, Mumbai-400021* .

[6] L. Fu. And J. Dai. , "A Speech Recognition Based on Quantum Neural Networks Trained by IPSO," International Conference on Artificial Intelligence and Computational Intelligence 2009.

[7] Lufen. "Research on Quantum Neural Networks Based Pattern recognition," [D].Nanjing University of Posts and Telecommunications, 2003.

[8] W. Ming-zhe, L. Xi and R. Li-mian, "Improved DTW Speech Recognition Algorithm Based on the MEL Frequency Cepstral Coefficients," Information and Communication Technology and Smart Grid, 2010.

[9] S. Singh and Dr. E.G. Rajan; "Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC," International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011.

[10] R. Yadav , D. Mandal , "Speaker Recognition using Particle Swarm Optimization (PSO)," IJECT Vol. 2, Issue 3, Sept. 2011.

[11] S. Selva Nidhyananthan, R. Shantha Selva Kumari and G. Jaffino, "Text-Independent Speaker Identification using Residual Feature Extraction Technique," CiiT International Journal of Digital Signal Processing, Vol 4, No 3, March 2012.

[12] S. Swamy and K.V Ramakrishnan, "AN EFFICIENT SPEECH RECOGNITION SYSTEM," Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, August 2013.

[13] L. Selvaraj and B. Ganesan , "Enhancing Speech Recognition Using Improved Particle Swarm Optimization Based Hidden Markov Model," Academic Editor: Thomas M. Deserno ,2014 .

[14] R. Bharti and P. Bansal, "Real Time Speaker Recognition System using MFCC and Vector Quantization Technique," International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 1, May 2015.

[15] H. Bae, H. Lee. And S. Lee. , "Voice Recognition Based on Adaptive MFCC and Deep Learning," 978-1-4673-8644-9/16/$31.00_c 2016 IEEE.

[16] B. Kanisha and G. Balarishnanan. , "Speech Recognition with Advanced Feature Extraction Methods Using Adaptive Particle Swarm Optimization," Indra Ganesan College of Engineering, Trichirappalli, India, 2016.

[17] Z. Hong. , "Speaker gender recognition system," university of OULU 2017.

[18] George R. Doddington, "Speaker Recognition- Identifying People by their Voices," Proceedings of the IEEE, VOL 73, NO. 11, November 1976.

[19] Joseph P. Campbell, JR, S. Member, "Speaker Recognition: A Tutorial", PROCEEDINGS OF THE IEEE, VOL. 85, NO. 9, SEPTEMBER 1997.

[20] A. Madan & D. Gupta, "Speech Feature Extraction and Classification," A Comparative Review. International Journal of computer applications, 2014.

[21] D.E. Sturim, W.M. Campbell, and D.A. Reynolds, "Classification Methods for Speaker Recognition," Springer-Verlag Berlin Heidelberg 2007.