**Engineering and Technology Journal**
**Journal homepage:** engtechjournal.org

# Speaker Recognition Systems in the Last Decade – A Survey

**Ahmed M. Ahmed [a*], Aliaa K. Hassan [b]**

[a] University of Technology, Baghdad, Iraq, 110172@student.uotechnology.edu.iq

[b] University of Technology, Baghdad, Iraq, 110018@uotechnology.edu.iq

*Corresponding author.

Submitted: 04/02/2020 Accepted: 11/07/2020 Published: 25/03/2021

K E Y W O R D S

Speaker Recognition, Speaker Identification, MFCC, and Feature Extraction.

A B S T R A C T

*Speaker Recognition Defined by the process of recognizing a person by his\her voice through specific features that extract from his\her voice signal. An Automatic Speaker recognition (ASP) is a biometric authentication system. In the last decade, many advances in the speaker recognition field have been attained, along with many techniques in feature extraction and modeling phases. In this paper, we present an overview of the most recent works in ASP technology. The study makes an effort to discuss several modeling ASP techniques like Gaussian Mixture Model GMM, Vector Quantization (VQ), and Clustering Algorithms. Also, several feature extraction techniques like Linear Predictive Coding (LPC) and Mel frequency cepstral coefficients (MFCC) are examined. Finally, as a result of this study, we found MFCC and GMM methods could be considered as the most successful techniques in the field of speaker recognition so far.*

**How to cite this article**: A. M. Majid and A. K. Hassan, "Speaker Recognition Systems in the Last Decade – A Survey," Engineering and Technology Journal, Vol. 39, Part B, No. 01, pp. 30-40, 2021.
DOI: https://doi.org/10.30684/etj.v39i1B.1589


## 1. INTRODUCTION

Speaker Recognition (SR) is an automated technique of identifying an individual on the basis on his\her voice signal, which is a biometric method like other biometrics such as fingerprint, Palm, Retina, Iris, and Face recognition. The main difference between Speaker Recognition and other biometrics is that Speaker Recognition can be considered as the only technology that processes acoustic information, in contrast with other methods, which usually use image information. Another significant difference is the capability to service with telephone equipment, and that would make it more broadly applicable to diversity settings. Also, other biometric techniques often require specific hardware to be able to work correctly [1].

ASR systems can be mainly classified into identification and verification because these two are the most widely used and traded technologies. Speaker verification (sometimes referred to as authentication) represent the process of verifying the claimed identity of an individual, whereas

Speaker Identification is the process of recognizing (identifying) a person from a set of many individuals. There are two various types of speaker identification, which are open-set and closed-set. In closed-set, the speaker-test is compared against all the speaker models within the database and return the speaker ID that produces the exclusive match, there is no rejection. On the other hand, Open-set can be considered as a closed set with verification task so it considered a complex problem than closed-set, sometimes speaker verification is taken into consideration as a unique case of open-set speaker identification [2].

ASR can be implemented in two methods, Text-dependent speaker recognition (TDSR) and Text-independent speaker recognition (TISR). In TDSR, the speaker uses a specific phrase that would be known to the system. On the other hand, in TISR, the speaker can use any phrase because the system does not have any stored phrase to compare with. Therefore, TISR is more challenging than TDSR. TDSR primarily applied to the speaker verification type. Whereas TISR primarily applied to the speaker identification type [2].

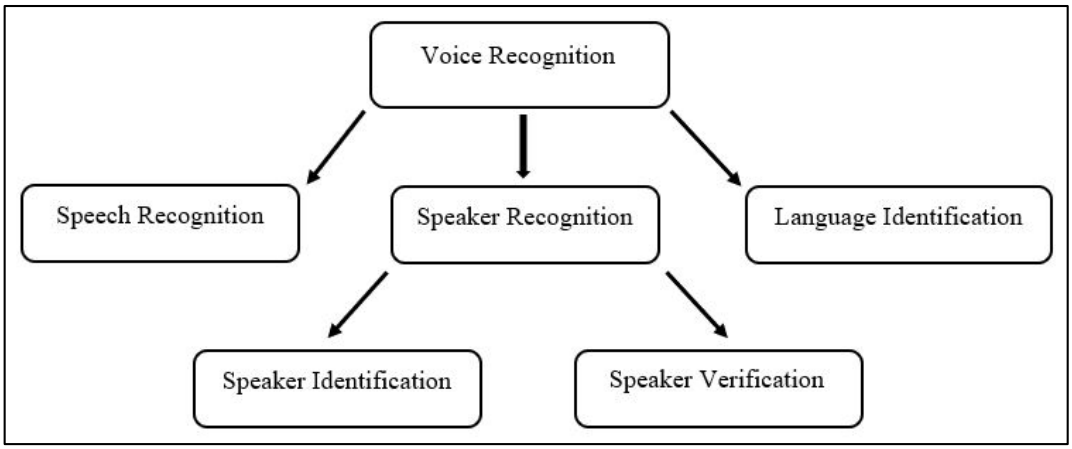


**Figure 1: Voice Recognition Categorization.**

All speaker recognition systems formed of two parts: feature extraction and feature matching. Feature extraction handles extracting some data from the voice signal of the speaker, whereas feature matching involves the comparison of the extracted characteristics with those already stored in the database [2].

There are two variation in the voice signal, which are Inter-speaker variability and Intra-speaker variability. Inter-speaker variability defines the variation in different person's voices. Whereas the Intra-speaker variability defines the variation in the same individual voice [3].

This work's main contribution consists of two points; the first is to make a comprehensive overview of the most recent advances and ideas in the field of speaker recognition technology. On the other hand, the second point is to clarify what techniques for speaker recognition systems in terms of feature extraction methodology and pattern matching method that provide the best outcomes for different speaker recognition systems.

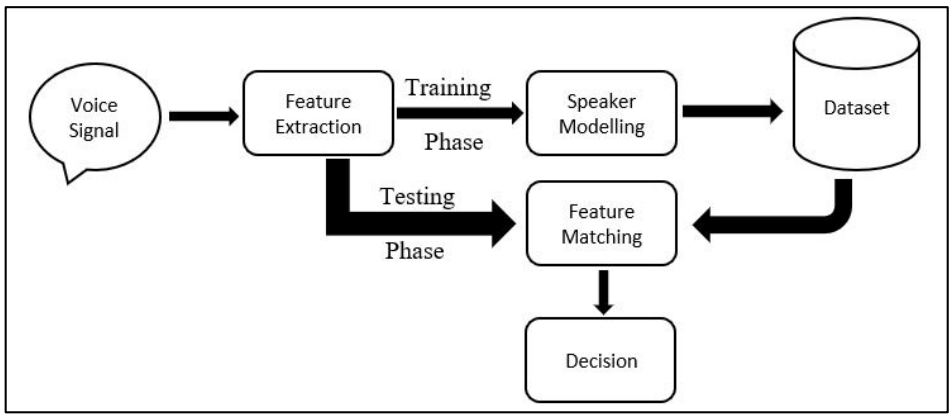


**Figure 2: Block diagram of speaker recognition system.**

The rest of this paper is arranged as follows: A summarized review of the most recent methods used in the ASR field is presented in Section 2. A review on the various feature extraction techniques are shown in Section 3. The most recent feature matching (speaker modeling) techniques are reviewed in section 4. Discussion in Section 5 and finally, the conclusion is presented in Section 6.

## 2. SPEAKER RECOGNITION DEVELOPMENT

The survey that we present in Table I, is the advancement of the speaker recognition system in the last decade (2010 - 2019), it gives us a brief overview of what has been done in the last decade in the area of speaker recognition technology. The titles of this Table's columns are organized as follows:

1) Author / Year
2) This refers to the researcher who developed the techniques and when the research is done.
3) Country
4) This refers to the country where the research was conducted.
5) System Type
6) Indicates whether the system is (Text-dependent or Text-independent) and whether it is (Identification or Verification).
7) Feature Matching (Modeling) Techniques
8) This refers to the speaker modeling techniques that are used in the research.
9) Features Extract. Method
10) This refers to the feature extraction methodology that are used in the research.
11) Dataset
12) This refers to the database type whether it is a popular dataset or private which have been conducted for the experiment.
13) Population (no. of speaker)
14) This refers to the number of people in the database used for the experiment.
15) Spoken Language
16) This refers to the language of the speaking people in the database being used.
17) Voice Type
18) This refers to the way the utterances were recorded.
19) Accuracy
20) It indicates how much that proposed system was accurate.

During the survey, it was realized that there are several areas where speaker recognition technology has been used. Areas with the most attention were access control, Telephone banking, Authentication, crime investigation, etc.

**TABLE I: Speaker Recognition progress in the last Decade.**

| Author / Year | Country | System Type | Feature Matching / Modeling Technique | Feature Extract. Method | Dataset | Number of speakers | Spoken Lang. | Voice Type | Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| Yun Lei et al./2010 | USA | Speaker Verification | GMM-UBM | MFCC | NIST SRE 2008 | 1543 | English | Telephone | **Error Rate**: 11.79%, 7.89% |
| Pawan K. Ajmera et al./2011 | India | Text-independent Speaker Identification | Nearest Neighbor Classifier | MFCC | TIMIT, SGGS, SGGS-2 | 630 151 36 | English, Hindi | Lab | 96.69% for TIMIT, 98.41% for SGGS |
| S. Sadiq et al./ 2011 | Turkey | Text-independent Speaker Identification | GMM | MFCC | TIMIT | 20 | English | Lab | 93%, 94% |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| C. Turner et al./2011 | USA | Text-dependent Speaker Identification | GMM | MFCC | TIMIT | 6 | English | Lab | **Error Rate:** 19% - 5% |
| Hesham Tol. / 2011 | Egypt | Text-dependent and text-independent Identification | CHMM | MFCC | Private Dataset | 10 | Arabic | Lab | 100% - 80% |
| Wei-Tyng Hong / 2012 | Taiwan | Text-independent Identification | HCRF-UBM | MFCC | MAT2000 | 300 | Mandarin | Lab | **Error Rate**: 7.2%, - 10.3% |
| Li Zhu et al./2012 | China | Speaker Identification | VQ | LPCC | Private Dataset | 20 | Mandarin | Lab | 89.33% - 91% - 94.67% |
| Ismail Shahin / 2013 | UAE | Text-independent and emotion-dependent identification | CSPHMM2s | LFPCs | OSD (Private Dataset) | 50 | English | Speech Acquisition Board | 81.50% |
| Fan-Zi Zeng et al./2013 | China | Speaker Verification | Hybrid (DFOA-SOM-PNN) | MFCC | Private Dataset | 20 | Mandarin | Lab | 99.57% |
| R. Ga. et al./2013 | Slovenia | Speaker state Recognition | HMM-UBM-MAP | MFCC | FAU AIBO – VINDAT | 51 children – 10 adults | Slovene - German | Standard Protocol | 71.5% - 70.9% |
| Anthony Larch. et al./2013 | UK | Text-dependent Speaker Verification | EBD (GMM-UBM), SCHMM | LFCC | MYLDEA | 30 | English | Microphone | **Error Rate**: 1.11%, 0.84% |
| Srikanth R Madi. / 2014 | India | Text-independent | PPCA-FA, | MFCC, MFS | NIST SRE 2010 | Unknown | English | Telephone Calls | **Error Rate**: MFCC (5.9-4.6-2.6) MFS (6.7-3.5-2.6) |
| Khaled Da. et al./2015 | Saudi Arabia | Speaker Identification and Verification | FWENN | Formant and Entropies | Private Dataset | 80 | Arabic | University Office | Verification (89.16%), Identification (90.09% - 82.5%) |
| Mansour Al Sula. et al./2016 | Saudi Arabia | Text-dependent and Text-independent | Diagonal GMM | MFCC, MDLF, MDLF-MA | KSU Database | 267 | Arabic | Room – Office - Cafeteria | 84%-25%-86% |
| S. Dey et al./2017 | Switzerland | Text-dependent Verification | GMM-UBM, I-vector | MFCC | RSR, Red-Dots | 190 females – 192 males | English | Standard Protocol | Error Rate: 0.18 |
| Rania M. G. et | Egypt | Text-independe | FHMM | WPFDF | Private Dataset | 100 | Arabic | Normal Office | 98.38% |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| al./2017 | | nt Verificatio n | | | | | | | |
| Suma Paulose et al./2017 | India | Text-independe nt closed-set Identificati on | GMM, I-vector | MFCC, IHC, (pitch and formants) | TIMIT | 100 | English | Lab | MFCC (GMM = 96.66%, I-vector = 91.33%), IHC (GMM = 82.33%, I-vector = 80.33%) |
| Nayana P.K et al./2017 | India | Text-independe nt closed-set Identificati on | GMM, I-vector | PNCC, RASTA PLP + Pitch and Formants | TIMIT | 100 | English | Lab | PNCC (GMM = 97.7%, I-vector = 90.7%), RASTA PLP (GMM = 89%, I-vector = 77%) |
| A. Antony et al./2018 | India | Text-dependent and Text-independe nt identificati on | ANN | MFCC, UMRT | Private Dataset | 15 | English | Microphone | Text-dependen t (97.91%), Text-independ ent (94.44%) |
| M.S. At. et al./2018 | India | Speaker Verificatio n from codec distorted speech | GMM-UBM + SVM | MPNCC + MFCC | TIMIT | 610 | English | Lab | **Error Rate**: 2.5% |
| Sit. A et al./2018 | India | Text-independe nt Closet-set Identificati on | (GMM-UBM), I-vector | MFCC, IHC | TIMIT | 600 | English | Lab | Four Experime nts: 94.22% - 80.8% - 66% - 44.22% |
| S. Hour. et al./2019 | Morocco | Text-independe nt Verificatio n | Similarity Measurem ents (Clusterin g) | MFCC | THUYG-20 SRE (Training ), FSCSR (Testing) | 371 - 536 | Uyghur – (6 differen t Langua ges) | Carbon Microphone | **Error Rate**: 0.77% - 0.32% |

## 3. FEATURE EXTRACTION TECHNIQUES

The purpose of the feature extraction method is to extract a condensed, effective set of parameters that reflect the acoustic impedance observed for subsequent use from the input speech signal. Feature extraction is the tool used to reduce the voice signal data aspect while maintaining the necessary information. Speech signal includes tons of information not all required to identify the speaker. Good features should be resilient against noise and distortion, should appear frequently and of course, in speech, should be easy to determine from voice sound, and should be difficult to mimic. The extracted characteristics can be classified into spectral, Spectro-temporal, speech source, short term, prosodic, etc. Short term spectral features are extracted from speech signals by dividing them into small frames of lengths of 20-30ms. [4].

In this section, we will look at different feature extraction techniques used by researches throughout the last decade.

***I. Mel Frequency Cepstral Coefficients (MFCC)***

MFCC is an audio extraction method that extracts speaker-specific parameters from the speech. (MFCC) is the most common and dominant method of extracting spectral characteristics for the speech by using the Fourier Transformed signal processing of the perceptually dependent Mel spaced filter bank [5]. According to Table 1, MFCC is very powerful, used by the most majority of the proposed systems, ideal for both Identification and Verification, and can be combined with other feature extraction techniques to provide better results.

***II. Linear Predictive Cepstral Coefficients (LPCC)***

In [6], weighted LPCC is proposed to improve the effectiveness of the feature parameter. LPCC is one of the predominant feature extraction techniques. The simple idea behind LPCC is to predict one sample of the speech at the current time as a linear combination of the previous samples [5]. The results approved that the weighted version of LPCC achieve a little improvement compared with traditional LPCC, and the recognition rate was (94.67%) for speaker identification.

***III. Log Frequency Power Coefficients (LEPCs)***

In [7], proposed a method for improving the recognition rate for the speaker identification system based on emotional states (emotion dependent). LEPC is a feature extraction technique that depicts the phonetic content of the speaker's utterance signal. The proposed technique achieved 81.5% recognition accuracy, and that may have a concern with the small size of the Dataset that was used (50 Speakers).

***IV. Linear Frequency cepstral Coefficients (LFCC)***

In [8], the LFCC feature extraction technique with an MYLDEA Dataset and three modeling techniques are proposed. LFCC is a feature extraction technique used in the field of speaker recognition. The recognition rate of this work was (98.99%), and that was acceptable in the area of text-dependent speaker verification.

***V. Mel Filter-bank Slope (MFS)***

MFS is a feature extraction that was rarely used in the previous decade for speaker recognition systems. Madikeri [9] used this technique as a trade-off with the MFCC technique. The experiments in that paper show that MFCC has slightly better results than MFS.

***VI. Formants and Entropies***

In [10], proposed a method used two feature extraction techniques to extract formants and entropies and these techniques were Power Spectrum Density (PSD) and Wavelet Packet (WP) respectively. They used these features with a proposed neural networks model called (FWENN) for speaker identification. Formants and Entropies are features extracted from the voice signal, specifically from the vowels rather than words or sentences like other features. The dataset used in this study contained 80 speakers and the recognition rates were (Verification (89.16%), Identification (90.09% - 82.5%)) for the Arabic language.

***VII. Multi Directional Local Feature (MDLF) and Multi Directional Local Feature with Moving Average (MDLF-MA)***

MDLF and MDLF-MA are feature extraction techniques that extract features in the time-frequency domain in different directions. In [11], proposed a method used these features combined with well-known MFCC to observe the speaker recognition performance based on Arabic phonemes. A good Dataset called KSU with a diagonal GMM for matching was not enough to achieve an impressive result. The recognition rates were (84%,82%,86%).

***VIII. Wavelet Packet Four-Directional Features (WPFDF)***

In [12], proposed a WPFDF which is a feature extraction technique for a text-independent speaker verification system based on Arabic utterances. WPFDF has subsequent merits. Firstly, they estimate four kinds of crucial information, that involve the acoustic proof of sounds of the sharp rising and fallings, spectral peaks within steady sounds, formant transitions along with voice

onset/offset. Secondly, they can diminish the variances resulted from the differences among speakers: the utterances and sentences; in addition to the speaking styles. For a speaker verification system, the results obtained from those experiments were fairly acceptable (98.38%).

### *IX. Inner Hair Cell Coefficients (IHC)*

IHCs are short-term features like MFCC, used for speaker recognition systems. In contrast with MFCC, IHC looks after the physiological variation in the auditory system of mammals. Paulose et al [4] used this technique with MFCC, pitch, and formants to observe the performance of several speaker modeling techniques when used with MFCC and IHC. The results show that MFCC is more quality than IHC, especially when combined with pitch and formants.

### *X. Power Normalized Cepstral Coefficients (PNCC) and Relative Spectral Perceptual Linear Prediction (RASTA PLP)*

PNCC is a modern feature extraction technique that works robustly even in noisy conditions. Whereas RASTA PLP is another feature extraction method that makes sure the generated signal to be less vulnerable to slow varying stimuli. Nayana et al [13] made a comparison between these features with different modeling methods for the text-independent Identification system. The experiments show that PNCC is much better than RASTA PLP.

### *XI. Modified Power Normalized Cepstral Coefficients (MPNCC)*

MPNCC represents a modified version of PNCC that used in Athulya et al [14] with the traditional MFCC for speaker verification technique from codec distorted speech. This experiment shows that MPNCC has more advantages over MFCC and PNCC in the situation of codec distortion speech, it also shows that PNCC in slightly better than MFCC in the same mentioned case. However, when a strident voice can be obtained, MFCC is superior.

### *XII. Unique Mapped Real Transform (UMRT)*

UMRT represents an extension of the transform technique (MRT), it is a signal processing technique that derived from Discrete Fourier Transform (DFT). Antony et al [15] used this technique with MFCC as a combination of feature extraction techniques for isolated word speaker identification. The results of the experiments show that when MFCC used with UMRT there is an improvement in the recognition rates as compared to MFCC only.

## 4. FEATURE MATCHING (SPEAKER MODELLING) TECHNIQUES

Feature Matching is a measure of the similarity between vectors with unknown features and reference models. Each signal model is constructed from the characteristics derived from the signal itself. The matching algorithm compares the received signals to the pattern of reference and indicates the distance. The distance is later used for identifying the unidentified speaker [7]. There are many types of models including GMM, Hidden Markov Models, and Vector Quantization (VQ), that can be used in speaker recognition. [16]. In this section, a general overview of the several feature matching techniques that have been used for speaker recognition systems in the last decade will be presented.

### *I. Gaussian Mixture Model (GMM)*

GMM is a probabilistic modeling technique that provides the probability distribution of multi-dimensional feature vectors extracted from the speaker's voice signal [13]. Statistically, GMM characterizes by three parameters (Mean, Covariance, and mixture weights). In [14], "GMM has shown to be the best for text-independent speaker recognition tasks".

As a whole, GMM can be described as the most well-known model technique for speaker recognition tasks, and the primary advantage of this method represents simplicity. GMM can merely be modified or combined with other techniques to improve the outcome of speaker recognition experiments. According to the previous works in the area of speaker recognition, GMM achieved good results. In [4], a comparison is made between GMM and i-vector techniques for text-independent identification. The results of the GMM were far better than the latter method.

***II. Continuous Hidden Markov Model (CHMM)***

In [7], proposed a modified version of the HMM (Hidden Markov Model) to carry out a text-independent speaker identification method. CHMM used Gaussian density functions and continuous observations to build the speaker model directly without any quantization. CHMM used with MFCC features for both text-dependent and text-independent. The recognition rate for the text-independent was 80%.

***III. Hidden Conditional Random Fields (HCRFs)***

In [17], proposed an approach for text-independent speaker identification, HCRF with UBM (Universal Background model) was used to model 300 speakers in the Mandarin Language from the MAT2000 dataset. HCRS is a speaker-recognition model technique based on conditional probability. The study investigates the performance of HCRF-UBM against other techniques GMM-UBM and HMM-UBM for identification. The outcomes of these experiments show that HCRF attained the smallest error rates among all the other techniques.

***IV. Vector Quantization (VQ)***

VQ is a digital signal processing method, commonly used for feature matching procedure in speaker recognition systems. Through an LBG algorithm (LINDE–BUZO–GRAY algorithm), In [6], VQ used to model 20 speakers with LPCC feature parameters. The study achieved somewhat acceptable results. However, the main drawback of this work is the limited size of the dataset.

***V. Second-Order Circular Suprasegmental Hidden Markov Models (CSPHMM2s)***

In [18], proposed CSPHMM2 which is a classifier technique for the speaker recognition system. This method is developed to improve the performance of the speaker identification that subjects to shouted talking situations. The study was performed on 50 speakers and utilized LFPCs features. The result of this experiment was moderate, the recognition rate he achieved was 81.5%.

***VI. Novel Hybrid Algorithm (DFOA-SOM-PNN)***

In [19], proposed a novel hybrid algorithm for speaker recognition in the Hybrid Algorithm. This algorithm used to enhance the performance of the Probabilistic Neural Network (PNN) in recognition. The proposed algorithm contains three sub-algorithms: 1) Self-Organizing Map (SOM) algorithm, a clustering algorithm used to cluster the characteristics parameters of the MFCC features, and to enhance the storage and the calculation time. 2) Double group Fruit Fly Optimization Algorithm (DFOA) to improve the PNN's smooth factor. 3) Probabilistic Neural Network (PNN) to allow the speaker to recognize. The results of this study for speaker verification were 99.57%, but the size of that used dataset was inadequate.

***VII. Hybrid PPCA-FA***

In [9], proposed A hybrid algorithm for text-independent speaker recognition. The proposed hybrid algorithm consists of Probabilistic Principal Component Analysis (PPCA) and a traditional I-vector method referred to as FA (Factor Analysis). The study used NIST SRE 2010 dataset and achieved good results, particularly with MFCC features.

***VIII. Feed-Forward and Probabilistic neural network (FWENN)***

In [10], proposed FWENN which is a classifier method for both speaker identification and speaker verification tasks. The proposed method applied two different classification techniques (Feedforward and Probabilistic Neural Networks). In [10], FWENN with feature extraction based on formants and entropies to model a database with 80 Arabic speakers. In the end, the outcomes of the identification and the verification tasks were (Verification: 89.16%, Identification: 90.09% - 82.5%).

***IX. Fuzzy Hidden Markov Model (FHMM)***

In [12], proposed a modeling-based speaker recognition technique called (FHMM) to decrease information loss and increasing the recognition rate of text-independent speaker verification. The study achieved good results by using this method combined with a proposed feature extraction technique termed as WPFDF.

### *X. Identity Vector (I-vector)*

I-vector is a modeling technique obtained from the GMM super vector. The I-vector based speaker recognition method has frequently been used for speaker recognition tasks in many previous works. The key principle of this technique is the ability to map the extracted features from every utterance into a region of dimensions very minimal than that of the total variability subspace [13].
Paulose et al [4] and Nayana et al [13] used this technique against GMM with various features (MFCC, IHC, PNCC, and RASTA PLP) to observe the performance of text-independent speaker identification with these two models. Both studies show GMM is much better than I-vector with all different features.

### *XI. Artificial Neural Networks (ANNs)*

In [15], proposed two ANN methods with a combination of UMRT and MFCC features to build an isolated word speaker identification system ANNs are information processing techniques that can be used to obtain patterns, knowledge, or models from a large amount of data. The obtained recognition rate form this study was: Text-dependent (97.91%), Text-independent (94.44%), but 15 speakers may be considered as an inadequate number for a dataset.

### *XII. Clustering Algorithms*

Clustering represents the process of dividing a group of objects into classes of similar objects. Several clustering algorithms have been used for speaker recognition tasks, and the crucial advantage when using a clustering technique instead of traditional stochastic methods is the cost in terms of time-consuming [14].

In [20], proposed a novel scoring technique based on similarity measurements for text-independent speaker verification. The study used two clustering algorithms (K-means and the nearest cluster algorithm) with MFCC features. The outcomes of this experiment achieved a slight improvement over the traditional GMM-UBM and I-vector methods which are considered as the state-of-the-art approaches for speaker recognition.

## 5. DISCUSSION

In this paper, various works are analyzed for examining the outcomes of different speaker recognition systems in the last decade. This survey can support us to develop an estimated solution for better results in future works. Various techniques are reviewed through the paper for every stage in the speaker recognition system structure (feature extraction, feature matching, and Datasets). There are many challenges the developers have to face to build a robust and powerful speaker recognition system. The most familiar challenges are inter-variation, intra-variation, mismatch channel, disguising of voice, background noise, etc. Although speaker recognition systems proved its ability in the area of biometrics, there are many factors like the dynamic behavior of the speech signal with the requirement of working in real-time still increase the complexity of the process.

Feature extraction represents a key stage in all speaker recognition systems. Many techniques used for this stage through all studies in the last decade. However, MFCC can be considered as the standard feature extraction method for speaker recognition. Most of the analyzed works in this survey used MFCC or a hybrid of MFCC with other techniques. Moreover, there are some studies made a comparison between this feature's technique with other techniques such as IHC and MFS, the outcomes show that MFCC superior the other mentioned methods.

In terms of Dataset, some works used a popular dataset such as TIMIT, and others used a private one for the study. The study shows it's very important to adapt the system with a dataset that contains a different utterance to make the system more robust against unexpected conditions. For the matching stage, there are extremely broad choices for a speaker recognition methodology. The matching technique could be a stochastic method, classifier technique, clustering algorithms, etc. Also, many of the previous works proposed many alternative techniques by either combining more than one method or modified a traditional one.

## 6. CONCLUSIONS

In this paper, we presented a comprehensive survey of speaker recognition systems in the last decade. Several concepts like Feature extraction techniques, modeling methods, challenges, etc. have been discussed. We have classified the modules and have shown many issues pertaining to the

speaker-recognition systems. Although many feature extraction techniques have been used throughout the last decade. MFCC still the standard feature extraction technique in speaker recognition systems. Frequently, MFCC provides good results and for some conditions and can be combined with other techniques to improve the performance of the system.

According to this survey, we can conclude that it's substantial to model the system with an adequate dataset in terms of the number of speakers and the diversity of syllables. Furthermore, the selection of the appropriate matching method depends on the problem at hand. The developer should concern many factors to adopt a valid method. But according to the survey, GMM modeling techniques is still the best and the state-of-the-art method for speaker recognition technologies.

## REFERENCES

**[1]** Z. Squib, N. Salam, R. P. Nair, N. Pandey, A. Joshi, A survey on automatic speaker recognition systems, Common. Computer. Inf. Sci., 123 (2010) 134–145. https://doi.org/10.1007/978-3-642-17641-8_18

**[2]** H. Beigi, Fundamentals of Speaker Recognition; Springer New York, NY, USA, 2011.

**[3]** N. Singh, A. Agrawal, The Development Of Speaker Recognition Technology, Int J Eng Adv Technol., 9 (2018) 8–16.

**[4]** S. Paulose, D. Mathew, A. Thomas, Performance Evaluation of Different Modeling Methods and Classifiers with MFCC and IHC Features for Speaker Recognition, Procedia Computer. Sci., 115 (2017) 55–62. https://doi.org/10.1016/j.procs.2017.09.076

**[5]** A. Sithara, A. Thomas, D. Mathew, Study of MFCC and IHC feature extraction methods with probabilistic acoustic models for speaker biometric applications, Procedia Computer. Sci., 143 (2018) 267–276. https://doi.org/10.1016/j.procs.2018.10.395

**[6]** L. Zhu, Q. Yang, Speaker Recognition System Based on weighted feature parameter, Phys. Procedia., 25 (2012) 1515–1522. https://doi.org/10.1016/j.phpro.2012.03.270

**[7]** H. Tolba, A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach, Alexandria Eng. J., 50 (2011) 43–47. https://doi.org/10.1016/j.aej.2011.01.007

**[8]** A. Larcher, J. Bonastre, J. S. D. Mason, Constrained temporal structure for text-dependent speaker verification, Digit. Signal Process., 23 (2013) 1910–1917. https://doi.org/10.1016/j.dsp.2013.07.007

**[9]** S. R. Madikeri, A fast and scalable hybrid FA / PPCA-based framework for speaker recognition, Digit. Signal Process., 32 (2014) 137-145. https://doi.org/10.1016/j.dsp.2014.05.012

**[10]** K. Daqrouq, T. A. Tutunji, Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers, Appl. Soft Comput. J., 27 (2015) 231–239. https://doi.org/10.1016/j.asoc.2014.11.016

**[11]** M. Alsulaiman, A. Mahmood, G. Muhammad, Speaker recognition based on Arabic phonemes, Speech Commun., 86 (2017) 42–51. https://doi.org/10.1016/j.specom.2016.11.004

**[12]** R. M. Ghoniem, K. Shaalan, ScienceDirect Science Direct A Novel Arabic Text-independent Speaker Verification System based on Fuzzy Hidden Markov Model, Procedia Computer. Sci., 117 (2017) 274–286. https://doi.org/10.1016/j.procs.2017.10.119

**[13]** P. K. Nayana, D. Mathew, A. Thomas, ScienceDirect Comparison of Text Independent Speaker Identification Systems using GMM and i-Vector Methods, Procedia Computer. Sci., 115 (2017) 47–54. https://doi.org/10.1016/j.procs.2017.09.075

**[14]** M. S. Athulya, P. S. Sathidevi, Speaker verification from codec distorted speech for forensic investigation through serial combination of classifiers, Digit. Investigation., 25 (2018) 70-77. https://doi.org/10.1016/j.diin.2018.03.005

**[15]** A. Antony, A. Antony, R. Gopikakumari, Speaker identification on combination of MFCC based features Speaker identification based on combination of MFCC and UMRT features," Procedia Computer. Sci., 143 (2018) 250–257. https://doi.org/10.1016/j.procs.2018.10.393

**[16]** S. Sadiç, M. B. Gülmezoğlu, Common vector approach and its combination with GMM for text-independent speaker recognition, Expert Syst. Appl., 38 (2011) 11394–11400. https://doi.org/10.1016/j.eswa.2011.03.009

**[17]** W. Hong, HCRF-UBM approach for text-independent speaker identification, Computer. Math. with Appl., 64 (2012) 1120–1127. https://doi.org/10.1016/j.camwa.2012.03.030

**[18]** I. Shahin, Speaker identification in emotional talking environments based on CSPHMM2s, Eng. Appl. Artif. Intell., 26 (2013) 1652–1659. https://doi.org/10.1016/j.engappai.2013.03.013

**[19]** F. Z. Zeng, H. Zhou, Speaker recognition based on a novel hybrid algorithm, Procedia Eng., 61 (2013) 220–226. https://doi.org/10.1016/j.proeng.2013.08.007

**[20]** S. Hourri, J. Kharroubi, A Novel Scoring Method Based on Distance Calculation for Similarity Measurement in Text-Independent Speaker Verification, Procedia Computer. Sci., 148 (2019) 256–265. https://doi.org/10.1016/j.procs.2019.01.068

**[21]** Y. Lei, J. H. L. Hansen, Mismatch modeling and compensation for robust speaker verification, Speech Commun., 53 (2011) 257–268,. https://doi.org/10.1016/j.specom.2010.09.006

**[22]** P. K. Ajmera, D. V. Jadhav, R. S. Holambe, Text-independent speaker identification using Radon and discrete cosine transforms-based features from speech spectrogram, Pattern Recognit., 44 (2011) 2749–2759. https://doi.org/10.1016/j.patcog.2011.04.009

**[23]** C. Turner, A. Joseph, M. Aksu, H. Langdon, The wavelet and Fourier transforms in feature extraction for text-dependent, filter bank-based speaker recognition," Procedia Computer. Sci., 6 (2011) 124–129. https://doi.org/10.1016/j.procs.2011.08.024

**[24]** R. Gajšek, F. Mihelič, S. Dobrišek, Speaker state recognition using an HMM-based feature extraction method, Computer. Speech Lang., 27 (2013) 135–150,. https://doi.org/10.1016/j.csl.2012.01.007

**[25]** S. Dey, P. Motlicek, S. Madikeri, M. Ferras, Template-matching for text-dependent speaker verification, Speech Common., 88 (2017) 96–105. https://doi.org/10.1016/j.specom.2017.01.009