Speaker Localization and Identification: A Survey of techniques

Rasha H. Ali, Dr. Mohammed Najm Abdullahand Dr. Buthainah F. Abed

Abstract

The speaker localization and identification simultaneously used vastly in smart and sensitive environments, meetings, conferences as well as in human-robot interactions. This paper attends a survey of the techniques that are used in speaker localization and identification and focused especially on one technique (using the RFID). The accuracy of results is very important in these systems. Speaker localization and identification can be accomplished in three stages: - preprocessing, feature extraction and lastly, classification or sometimes called decision-making stage. Therefore, this paper introduces some types of audio databases that are used in speaker localization and identification systems, the techniques that are used in feature extraction stage with the comparison between them, techniques that are used in classification stage and comparison between them. The technique of fusion (feature fusion or fusion decision-making) gives the improvement in the result and increase the accuracy.

***Index term*–Speaker localization, Speaker identification, feature extraction, classification, fusion technique**

I. INTRODUCTION

Contemporary revolution in smart environments requires real-time monitoring of interaction between human and computer or between humans. In the situation of many persons are present in a single room, the localization andidentification every person simultaneously is very important [1],[2], it used in numerous applications such as in smart hospital rooms, smart homes, interaction of human-robot, interactions between robots and in all applications that are required localize and identity the one person or multi persons with high challenges and high accuracy specially in now days that the mobile phone become very famous to use in recording or using as microphone because most of the recent mobile phone technologies contain accelerometers and magnetometers attached to them[3].

Speaker localization is an important techniques for the ability of robot to detect the directions of talkers for expression of the interest in the conversation by speaker localization systems [4]. Whereas the system of human auditory has abilities to realize and localize a target source in complex multi-source scenarios, so it remained the challenging task for algorithms[5].

Speaker recognition contains two methodologies, firstly Speaker identification(SI) and secondly speaker verification(SV). SI is determining the identity of unknown speaker depend on his/her utterance properties. Speaker Identification(SI) is used in various application such as shopping using the telephone, voice dialing, mail of voice, and the services for access to database[6].

Overall, this paper offers the types of techniques for speaker localization and identification in section II. Section III presents some works those related to speaker localization and identification. Section IV presents the types of databases, The some techniques that are used in speaker localization were covered in section V. Section VI offers the stages for speaker identification. The types of feature extraction were covered in section VII. The feature extraction techniques were displayed in section VIII. Section IX displays the techniques of classification and conclusion was displayed in section X.

II.TECHNIQUES OF SPEAKER LOCALIZATION AND IDENTIFICATION

There are different resources of techniques for speaker localization and identification systems, some of these techniques depend on Radio-frequency Identification (RFID)**,** the other techniquesdepend on multiple modalities and others depend on camera video. The table I. displays the techniques of the localization and identification of speaker[7].

II. RELATED WORK

There is prior work relatedto localize and identify the speaker, some of them were using audio files (one microphone or microphone arrays), audio/ visual files or were using sensor/ audio and camera. In [11] speaker localization and identification in smart environments using multimodal information had been proposed, used the features of audio and visual for motion detection, person tracking and faceidentification. The result showed the matches percentage was 83.8% when using acoustic and visual features in identification and tracking while the matches percentage for the visual feature was 81.4%.

In[12] the speaker localization and identification wasimplemented in two stages:- the first was applied on YOHO database for identifying the speaker depend on MFCC and RMFCC and IAIF as feature extraction and GMM technique for training, the misclassification rate in this model was 10.13% for MFCC, misclassification for RMFCC was 30.96% and for IAIF was 62.04%. While the second stage was localizing the identity speaker using GCC with phase transform for estimating the differences between the times of speech signals, the misclassification rate was 24.67%.

Table I. The techniques of speaker localization and identification

The name of technique	Types	Approaches
Radio- frequency Identification (RFID)	Use one microphone or microphone arrays to extract the information for localization and identification of person or persons[8].	Using the differences of shifting times between signals, difference of arrival to estimate the location.
Camera Video	Use images from camera to detect the location of person or persons[9].	Using Probabilistic multi-hypothesis for tracking or tracking using hypothesis heuristicor both for more accuracy.
Multiple Modalities	Using sensors/video/audio as integrated system to capture every interaction between robot and human in different areas[10].	Using Audio- Visual signals by arrays of microphone and arrays of camera, use multi techniques of classification and clustering.

In Reference[13] the identification and localization were achieved simultaneously for mobile speakers depend on binaural signals from the robot in noisy and echoic environments. For localization stage, the method of the differences for interaural level (ILD) was used to implement it, the identification stage was implemented using equivalent rectangular bandwidth frequency cepstral coefficients (ERBFCC) method for feature extraction and ANN using for classification. The results offered good result when used in real environments with the presence the noise and echo.

IV. SPEAKER LOCALIZATION AND IDENTIFICATION DATABASES

There are many types of databases depend on some features such as recording protocol, the population of participating subjects, the language of recording device**,** type of oral statement, and the intended use, etc. Also, there are many types of databases, some of them are for meeting, conference, broadcasting news and conversation call, such as(TIMIT, ELSDSR, YOHO, Noisex, Santa Brabara corpus of spoken English)[14].

Also, select the database to depend on the type of speaker localization and identification techniques (Audio, Camera/Audio, and Camera/Audio/Sensor). weconcentrate in this paper on some databases that are used for the audio file.

Table II shows the various types of audio databases.

V. SPEAKER LOCALIZATION TECHNIQUES

Speaker localization depends on either one microphone or microphone arrays, the following some techniques that are using for speaker localization[16]. Such as time delay estimation (ITD), Direction of arrival estimation (DOA) and functions for head-related transfer (HRTFs) [17].

1. Inter-aural Time Difference (ITD)

Also known as Time-Difference Of Arrival (TDA or TDOA) or Inter-aural Phase Difference (IPD). It is the most popular technique for localizing the source of sound[18]. Since it depends on the shifting between two signals[19].

2. Direction of arrival (DOA)

This technique of localization depends on the assessment of the direction of access to information retrieval [20].

Table II.Databases types

Type	Purpose	No. of audio file
TIMIT	Using data for speech for the acquisition of acoustic- phonetic knowledge, and for the development and evaluation of automatic speech and speaker recognition systems[15].	630 voice messages for speakers, which (438 M/192 F) and each speaker reads 10 different sentences
YOHO	For speaker verification in cases of text-dependent in applications for secure access[14].	138 speech messages for speakers(106 M/32 F).
ELSDSR	For speaker (identification, localization and verification) and in speech recognition system[14].	The number of speakers is 22 (12M/ 10F), while the ages of the speakers between(24 – 63).

It used to detect the direction of different waves and electromagnetic sources from the number of receiving antennas that constitute a sensor array. DOA used in different applications like wireless communications, sonar, radar,.., etc.[20]. Figure I shows the two categories of DOA algorithms [21],[22].

3. Head-Related Transfer Functions (HRTF)

The transfer function of Head-related depends on capturing the path of acoustic transmission to the external ear from an acoustic source. The HRTF technique can be used in free-field propagation. The HRTF can be achieved in the domain of time, or implement related to the impulse response(HRIR). The HRIR also was known as the impulse response for binaural room (BRIRs)[17].

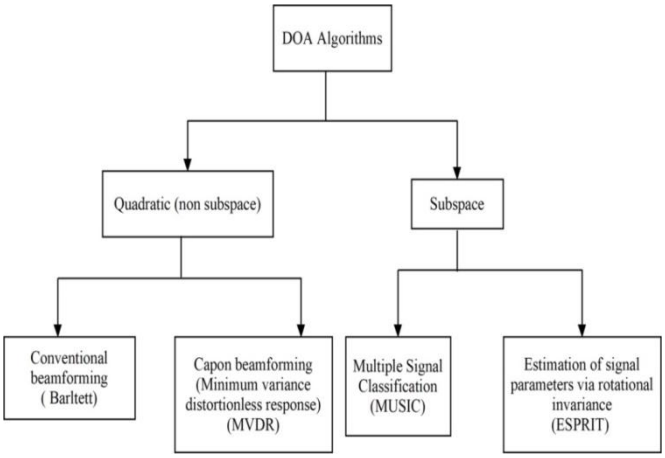

Fig.I. DOA Algorithms

removes the noise or background noise from signals using different techniques.

2. *Feature extraction:*- this step for extracting the acoustic features from signals of speakers to construct set of feature vectors for use in training and testing phases.
3. *Speaker database* :- which stored the feature vectors.
4. *Decision-making*:- this step for the final decision about the identity of the speaker by comparing unknown feature vectors to all models in the database and selecting the best matching model. Fig.II. shows speaker identification stages[23].

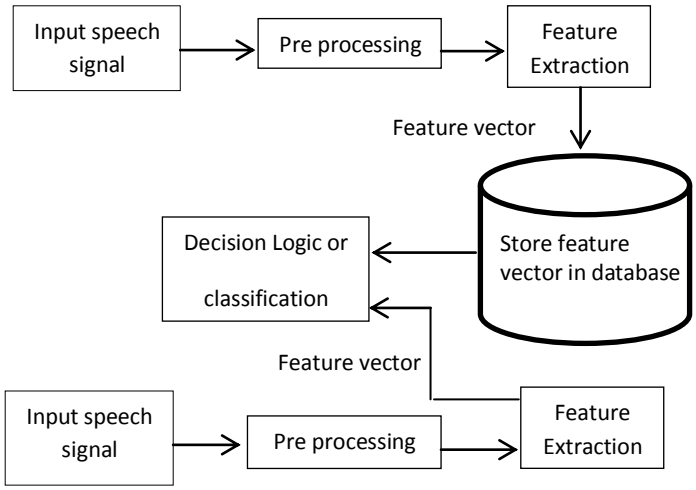

Fig.II. Speaker Identification stages

VI. SPEAKER IDENTIFICATION STAGES

The speaker identification contains the following stages:-

1. *Front-end processing:* - this step related to the "signal processing" part, which converts the continuous signals into discrete signals,

VII. Features types

The features in the stage of feature extraction have various types, such as voice source features, the features of spectral for short-term, the features of spectro-temporal, the features of high-level and prosodic features. The comparison between these features shown in the table III [24].

Table III. Features types

The type of feature	*The time of each type*	*The robustness*
Short-term spectral features	The features of short-term depend on short frames computation. The times of frames is about 20–30 ms in duration	Good, depend on physical characteristic
Voice source features	The features depend on voice feature by recognizing the source of voice (glottal flow).	Good, depend on physiological properties
Spectro-temporal and prosodic features	The features depend on spectral and prosodic features by extracting the spanning over hundreds of milliseconds[24].	more robust
High-level features	The features depend on the high features by extracting the characteristics of speakers for conversation-level.	more robust and gives good performance

VIII. FEATURE EXTRACTION TECHNIQUES

There are different techniques are using to construct feature vector, some techniques are processing in the time domain and other techniques are processing in the frequency domain. The following some technique which is used in feature extraction stage:-

1.Linear Predictive Coding (LPC)

The LPC construes the signal of speech by estimating the formants removing their effects from the speech signal[25]. The LPC is depended on linear prediction model by computing the minimizing the sum of the squared difference between actual speech samples and the linearly predicted sample. Fig. III. shows LPC structure[26].

2. Mel Frequency Cepstral Coefficients (MFCC)

This technique depends on ear scale of human by using the scale of Mel, based on the frequency domain [27]. It is representing the real cepstral of the short time of signal after computing the windowing technique, the transform of Fast Fourier (FFT) was used to transform into the domain of frequency. The nonlinear frequency scale which convergent to the behavior of the auditory system was applied for getting the differences from the real cepstral[28]

3. Perceptual Linear Prediction (PLP)

Perceptual linear predictive analysis (PLP) was suggested by HynekHermansky in 1989[27]. the PLP technique utilizes the psychophysics of system for hearing to used three concepts. These three connotations are (Intensity loudness power law, Equal loudness curve, and The critical-band spectral resolution). The PLP and LPC techniques are estimating the spectrum of short-term power for the signal of speech by using the all-pole model[28][29]. The properties of the previous techniques that depended on some criteria shown in the table IV.

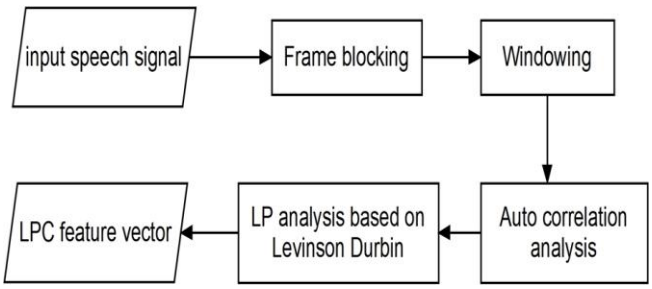

Fig.III. LPC structure

Table IV. The properties of LPC, MFCC and PLP techniques

criteria	MFCC	LPC	PLP
Application	Used in systems of voice recognition for security purposes	Used for overcoming the constraints on software that are associated with speech recognition system.	Used in systems of telephone connections to reduce the effects of differences in the frequency response.
Usage	To analyze a cepstrum of speech signal based on the spectral smoothing.	For estimating the formant to analyze the speech signal.	For estimating short term spectrum to analyze the speech signal.
correlation	Low correlation between coefficients	High correlation between features components.	Low correlation.
accuracy	Good in discrimination	Not sufficient for representation.	Not sufficient for representation.
Linearity	Not linear.	Linear scales	Linear scales
mathematically	well in discrimination.	Straightforward to implement and accurate in mathematic	straightforward to implement and accurate in mathematic

IX. DECISION MAKING (CLASSIFICATION)

Also, a known as speaker modeling. The vectors of the feature that are taken away from the utterances of speakers. These vectors of the feature are stored and trained in the database of the system for classification[30], [31]. In general, the models of classification stage which categorized into two models

- Nonparametric models (template Models) which training and test feature vectors are compared directly between them.
- Parametric models (stochastic models) which depend on training the feature vector [32]. The following some techniques that can be used for Speaker Modeling.

1. Dynamic Time Wrapping (DTW)

It a non-statistical tool that has been successfully applied in speaker recognition task, it is simple and effective technique[33],[34].It a dynamic programming technique based on separating the whole problem into a tiny number of procedures and each one used local distance measures to get a decision about each procedure. The summing of these smaller decisions is the overall decision[35].

2. Vector Quantization (VQ)

It a method of mapping vectors into a restricted number of parts from a vast vector space, each region is a cluster and representing each region by the center of that region called a code word. A codebook is the aggregate of whole code words. Hence, if there are multiple users so should be represented with multiple codebooks each codebook represents the identical speaker[36]. It a technique of unsupervised learning used in many application such as compression of data, encoding and for clustering the data[37],[38].

3. Support Vector Machines (SVM)

It a binary classification that determines the optimal linear decision surface based on the concept of minimization for structural risk instead of using the training data for modeling the probability distribution[39]. The surface of decision is a weighted collection of elements of a training set. The elements of this technique called (support vectors) which describe the border between two classes[40]. Figure VI. shows SVM structure.

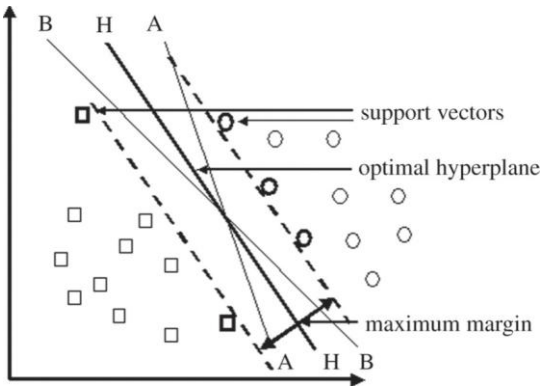

Fig. VI. SVM structure

4.Gaussian Mixture Model (GMM)-Universal Background Model (UBM)

It the stochastic model[45]. It used It used the maximum posteriori algorithm (MAP) or the algorithm of Expectation-Maximization (EM) for esteeming the parameters of GMM from training data[46]. UBM is also called as GMM when there are the massive set of speakers .UBM is used to compare the person's independent feature properties versus person specific feature model during decision of rejection or acceptance[47].

5. Hidden Markov model (HMM)

It a stochastic model and a powerful mathematical method used in the modeling the time series by estimating the state and parameter[41]. It depends on model of Markov, which is an automaton of finite state. This sequence of a state is called a chain of Markov, it depended on the Markov chains which used probability theory for modeling a sequence of events in time[42].

6. Artificial Neural Network (ANN)

It is powerful discrimination models, which used in the processing of digital signal when statistical distributions are not previously recognized[41]. The ANN comes in different types such as Multi-layer network (MLP) and network of Deep Neural(DNN). The uses of ANN comes in different stages of speaker localization and identification such as ANN that used either in classification or in feature extraction, while the features that extracted by ANN is called bottleneck features[42, 43].

In References[44]-[46] show that the ANN especially DNN was given good result and improve in performance in speaker or speech recognition than the other techniques. The comparison between the speaker modeling techniques shown in table V.

7. Fusion technique

Data Fusion means the information is combining from a different source of a directory. [47]. There are two types of fusion:- decision fusion and feature fusion[48]. The information in the fusion technique can be computed independently in different techniques such as Fuzzy integral approach, Voting/ranking methods, Dempster–Shafer approach and Log opinion pool[49]. Also, in the fusion decision, different classifiers which used for the same features[50]. In [51-53] the results had shown that the system which is used fusion technique either in the extraction of feature stage or decision stage, gives good performance than used one classifier system or feature alone.

Table V. Comparison between the speaker modeling techniques

Criteria	*DTW*	*VQ*	*SVM*	*GMM*	*HMM*	*ANN*
Mathematically	Not statistical and deterministic	Not statistical and deterministic	statistical and deterministic	statistical and deterministic	statistical and deterministic	deterministic
Linearity	Not linear	linear	linear	linear	linear	Not linear
Discriminatively	Not Discriminative	Not Discriminative	Discriminative	Discriminative	Not Discriminative	Discriminative
Usage	Used for calculating the symmetry for two sequences which different in speed or time.	Used for Reducing the size of memory that wanted to represent the signals.	Used for classification in high dimensional patterns.	For making a decision of acceptance or rejection in speaker, speech recognition systems.	Used for processing the signals by modeling the time distribution and used for estimating the parameter.	Used for classification the patterns Because its ability to self-learning and it adaptation in new environments.
Accuracy	Good	Good	Robust and excellent Classifier	Good	Good and simple to adaptation.	Robust and good Classifier
Length of inputs	Variable	Variable	Fixed length of inputs.	Variable	Length of inputs can be variable	Variable

X. CONCLUSION

This paper has displayed a survey of the types of techniques for localization and identification of speaker but especially audio files (RFID with one microphone or microphone arrays), types of audio databases for localization and identification of the speaker. Because using this technique in many application such as meetings and conference so the increase of accuracy of that system is very important. Wherefore, we displayed a survey for feature extraction techniques and decision-making techniques, with the comparison between the techniques of the feature extraction stage and between the techniques of the classification stage.

The results of researches showed amelioration in performance when using the neural network and the fusion technique. Also, there was the improvement in accuracy of speaker identification system when used feature fusion technique and when using fusion decision-making technique.

Reference

[1] Cucchiara, R., et al. "Mutual calibration of camera motes and RFIDs for people localization and identification", *in Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras.* 2010.

[2] Teixeira, T., D. Jung, and A. Savvides,"Tasking networked cctv cameras and mobile phones to identify and localize multiple people",*in Proceedings of the 12th*

ACM international conference on Ubiquitous computing, 2010.

[3] Bernardin, K. and R. Stiefelhagen, "Audio-visual multi-person tracking and identification for smart environments",*in Proceedings of the 15th ACM international conference on Multimedia*, 2007.

[4] Blauert, J. and J. Braasch," Binaural signal processing. in Digital Signal Processing (DSP)", *International Conference on, IEEE*,17th, 2011

[5] May, T., S. van de Par, and A. Kohlrausch, "Simultaneous localization and identification of speakers in noisy and reverberant environments" *in Proc. FORUM ACUSTICUM*, 2011.

[6] Campbell, J.P., "Speaker recognition: A tutorial", *Proceedings of the IEEE*, vol.85, no.9, pp. 1437-1462, 1997.

[7] Ferdous, S., K. Vyas, and F. Makedon,"A survey on multi person identification and localization",*in Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, 2012.

[8] Hahnel, D., et al,"Mapping and localization with RFID technology. in Robotics and Automation", *Proceedings. ICRA'04, IEEE International Conference*, 2004.

[9] Focken, D. and R. Stiefelhagen,"Towards vision-based 3-d people tracking in a smart room. in Multimodal Interfaces",*Proceedings. Fourth IEEE International Conference on*, 2002.

[10] Trivedi, M.M., K.S. Huang, and I. Mikic, "Dynamic context capture and distributed video arrays for intelligent spaces",*IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol.35, no.1: pp. 145-163, 2005.

[11] Salah, A.A., et al., "Multimodal identification and localization of users in a smart environment",*Journal on Multimodal User Interfaces*, vol.**2**, no.2, pp.75-91, 2008.

[12] Tómasson, H., "Speaker localization and identification", 2012.

[13] Youssef, K., K. Itoyama, and K. Yoshii, "Simultaneous Identification and Localization of Still and Mobile Speakers Based on Binaural Robot Audition", *JRM*,vol.**29**, no.1, p. 59-71, 2017.

[14] Feng, L. and L.K. Hansen, "A new database for speaker recognition", 2005.

[15] Garofolo, J.S., et al., "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM", *NIST speech disc 1-1.1. NASA STI/Recon technical report*, pp.93, 1993.

[16] Yang, Z., et al., "Sparse methods for direction-of-arrival estimation",*arXiv preprint* , 2016.

[17] Dmochowski, J.P. and J. Benesty, "Steered beamforming approaches for acoustic source localization",*Speech Processing in Modern Communication*, pp. 307-337, 2010.

[18] Lenz, C., "Localization of Sound Sources, Studies on Mechatronics", PhD Thesis, Autonomous Systems Lab, *Swiss Federal Institute of Technology Press*, 2009

[19] Bhadkamkar, N. and B. Fowler, "A sound localization system based on biological analogy", in Neural Networks, *IEEE International Conference on*, 1993.

[20] Lavate, T.B., V. Kokate, and A. Sapkal.,"Performance analysis of MUSIC and ESPRIT DOA estimation algorithms for adaptive array smart antenna in mobile communication", in Computer and Network Technology (ICCNT),*IEEE Second International Conference on*, 2010.

[21] Varade, S.W. and K.D. Kulat.,"Robust algorithms for DOA estimation and adaptive beamforming for smart antenna application".,*in Emerging Trends in Engineering and Technology (ICETET), 2nd International Conference on, IEEE*., 2009.

[22] Rao, B.D. and K.S. Hari, "Performance analysis of root-MUSIC",*IEEE Transactions on Acoustics, Speech, and Signal Processing,*. vol.37, no.12, pp. 1939-1949, 1989.

[23] Tiwari, V., "MFCC and its applications in speaker recognition",*International journal on emerging technologies*, vol.1, no.1, pp. 19-22, 2010.

[24] Doddington, G.R.,"Speaker recognition based on idiolectal differences between speakers. in Interspeech", 2001.

[25] Furui, S., "Speaker-independent isolated word recognition using dynamic features of speech spectrum",*IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.34, no.1, pp. 52-59, 1986.

[26] Nijhawan, G. and M. Soni, "A comparative study of two different neural models for speaker recognition systems",*International Journal of Innovative Technology and Exploring Engineering*, pp.2278-3075, 2012.

[27] Jamaati, M., H. Marvi, and M. Lankarany, "Vowels recognition using mellin transform and plp-based feature extraction",*Journal of*

the Acoustical Society of America, vol.123, no.5, pp. 3177, 2008.

[28] Gunawan, W. and M. Hasegawa-Johnson, "PLP coefficients can be quantized at 400 bps. in Acoustics", *Speech, and Signal Processing, Proceedings.(ICASSP'01), 2001 IEEE International Conference on*, 2001.

[29] Hermansky, H., "Perceptual linear predictive (plp) analysis of speech",vol.87, no.4, April, 1990.

[30] BenZeghiba, M.F. and H. Bourlard, "On the combination of speech and speaker recognition", *IDIAP*, 2003.

[31] Heck, L.P. and D. Genoud.,"Combining speaker and speech recognition systems",*in Seventh International Conference on Spoken Language Processing*. 2002.

[32] Kinnunen, T. and H. Li, "An overview of text-independent speaker recognition: From features to supervectors",*Speech communication*, vol.52, no.1, pp. 12-40, 2010.

[33] Oshaughnessy, D., "Speaker recognition",*IEEE ASSP Magazine*, pp. 4-17, 1986.

[34] Furui, S., "Cepstral analysis technique for automatic speaker verification",*IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.29, no.2, pp. 254-272, 1981.

[35] Amin, T.B. and I. Mahmood, "Speech recognition using dynamic time warping. in Advances in Space Technologies", *ICAST, 2nd International IEEE Conference on*. 2008.

[36] Ch, P., "Text dependent speaker recognition using MFCC and LBG VQ", 2007.

[37] Kekre, H. and T.K. Sarode, "Speech data compression using vector quantization",*WASET International Journal of Computer and Information Science and Engineering (IJCISE)*,vol.**2, no.**4, pp. 251-254, 2008.

[38] Banerjee, A., et al., "Clustering with Bregman divergences",*Journal of machine learning research*, vol.6, pp. 1705-1749, Oct2005.

[39] Vapnik, V.N. and V. Vapnik, "Statistical learning theory", Wiley New York, 1998.

[40] Campbell, W.M., et al., "Support vector machines for speaker and language recognition",*Computer Speech & Language*, vol.20, no.2, pp. 210-229, 2006.

[41] Lin, Q., et al.,"Speaker identification in teleconferencing environments using microphone arrays and neural networks"., *in Automatic Speaker Recognition, Identification and Verification*. 1994.

[42] Yamada, T., L. Wang, and A. Kai.,"Improvement of distant-talking speaker identification using bottleneck features of DNN",*in Interspeech*. 2013.

[43] Sainath, T.N., et al.,"Learning filter banks within a deep neural network framework",*in Automatic Speech Recognition and Understanding (ASRU), IEEE Workshop on*. 2013.

[44] Mohamed, A.-r., G.E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks",*IEEE Transactions on Audio, Speech, and Language Processing*, vol.20, no.1, pp. 14-22, 2012.

[45] Seide, F., G. Li, and D. Yu,"Conversational speech transcription using context-dependent deep neural networks", *in Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[46] Kingsbury, B., T.N. Sainath, and H. Soltau,"Scalable minimum Bayes risk training of deep neural network acoustic models using distributed Hessian-free optimization",*in Thirteenth Annual Conference of the International Speech Communication Association*. 2012.

[47] Ramachandran, R.P., et al., "Speaker recognition—general classifier approaches and data fusion methods", *Pattern Recognition*, vol. **35**, no.12, pp. 2801-2821, 2002.

[48] Nematollahi, M.A. and S.A.R. Al-Haddad, "Distant speaker recognition: an overview", *International Journal of Humanoid Robotics*, vol.13, no.2, 2016.

[49] Wang, L., N. Kitaoka, and S. Nakagawa, "Robust distant speaker recognition based on position-dependent CMN by combining speaker-specific GMM with speaker-adapted HMM",*Speech communication*, vol.49, no.6, pp. 501-513, 2007.

[50] Brummer, N., et al., "Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006", *IEEE Transactions on Audio, Speech, and Language Processing*, vol.**15**, no.7, pp. 2072-2084, 2007.

[51] Ding, J., C.-T. Yen, and D.-C. Ou, "A method to integrate GMM, SVM and DTW for speaker recognition", *International Journal of Engineering and Technology Innovation*, vol.**4**, no.1, pp. 38-47, 2014.

[52] Kinnunen, T., V. Hautamäki, and P. Fränti,"On the fusion of dissimilarity-based classifiers for speaker identification", *in INTERSPEECH*, 2003.

[53] Sree, S.S. and D.N. Radha, "A survey on fusion techniques for multimodal biometric identification",*International Journal of Innovative Research in Computer and Communication Engineering*, vol.2, no. 12, 2014.

AUTHORS

Rasha H. Ali :[Email: ha170642@gmail.com], Lecturer in University of Baghdad- College of Education for Women- Computer department and Ph.D. Student in ICCI.

Dr. Mohammed Najm Abdullah:
[mustafamna@yahoo.com], Asst. Prof. Lecturer in Iraqi Commission for Computers and Informatics (ICCI)- Information Institute for Postgraduate Studies

Dr. Buthainah F. Abed: [buthynna@yahoo.com], Asst. Prof. University of Information Technology and Communications