# STATE OF THE ART ON THE CURRENT RESEARCH LINES IN SPEAKER RECOGNITION BASED ON CLUSTERING METHODS

Z. Hamadache
USTHB University

***Abstract—*** The objective of this overview is to summarize some of the well-known algorithms already studied and tested for the task of speaker recognition during the recent years. First, we give an overview of speaker recognition, then we present the development and understanding of its state of the art by highlighting the contribution from the latest developed techniques in general, followed by the inclusion of the state of the art part in speaker recognition based on clustering techniques. Again, a special emphasis on the current research lines is given in order to know the new approaches of speaker recognition. Thus, an introduction on speaker recognition and a summary on the state of the art related to clustering methods are offered and discussed.

***Keywords***— Speaker recognition, State of the art, Clustering methods, Latest research.

## I. Introduction

Speech is the main way of communication between human beings. From speech signal, we can characterize many important characteristics about the speaker: the nature (sympathetic, respectful, etc…), the language, the approximate age (teenager, young, and old), the gender (male or female), the emotional state, the origin, the background (culture), etc... Speaker recognition (SR) is a biometric modality to recognize the individual who is speaking from a speech utterance, also called voice recognition regarding the personal basis of person's voice information, while speech recognition is not a biometric modality, but a method to recognize words as they are articulated. The following terms: “Speaker” and “Speech” recognition belong to the general term “Speaker recognition”, but, they are not the same even if "Voice recognition" can be used for both. Moreover, “speaker verification” (also called speaker detection or speaker authentication) and “speaker identification” are completely different. The differences between Speaker recognition and Speech recognition are shown in details in (Fig 1).

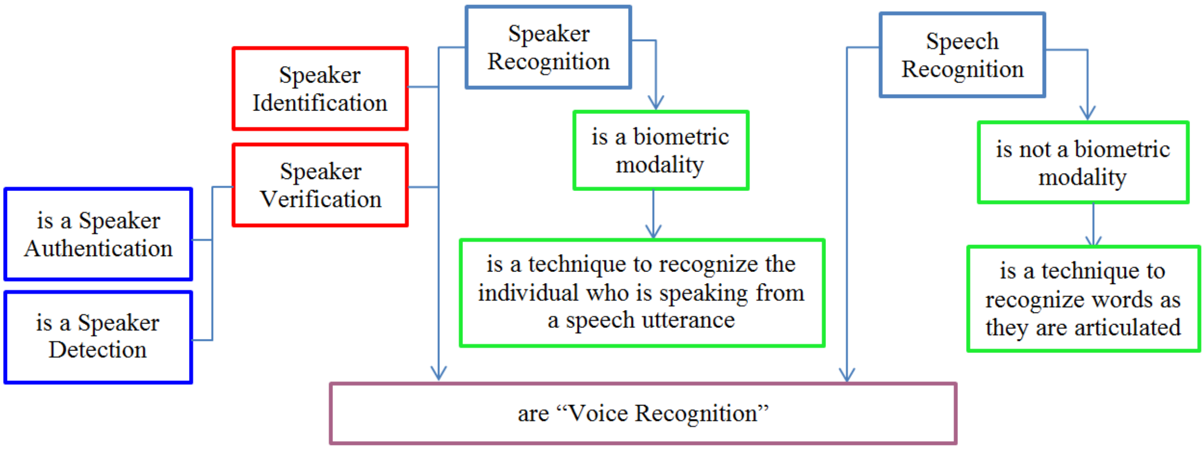
Speaker Identification
Speaker Verification
is a Speaker Authentication
is a Speaker Detection
Speaker Recognition
is a biometric modality
is a technique to recognize the individual who is speaking from a speech utterance
Speech Recognition
is not a biometric modality
is a technique to recognize words as they are articulated
are “Voice Recognition”

**Figure 1**: Differences between Speaker recognition and Speech recognition

The Speaker recognition (SR) performance has increased in recent years, which makes its application very important, mainly in security systems (e.g. tracking criminals). SR can be roughly grouped into two classes: 1) Conventional and 2) Text conversion or speech modalities. The Conventional group comprises: 1) Speaker identification and 2) Speaker verification (also called Speaker Detection or Speaker Authentication). In addition, Speaker identification is also categorized into two types associated to known voices: 1) Open set Identification and 2) Closed set Identification. While, the Text conversion or speech modalities group comprises: 1) Text-dependent recognition (also called constrained mode) and 2) Text-independent recognition (also called unconstrained mode). These classes are described in (Fig 2).

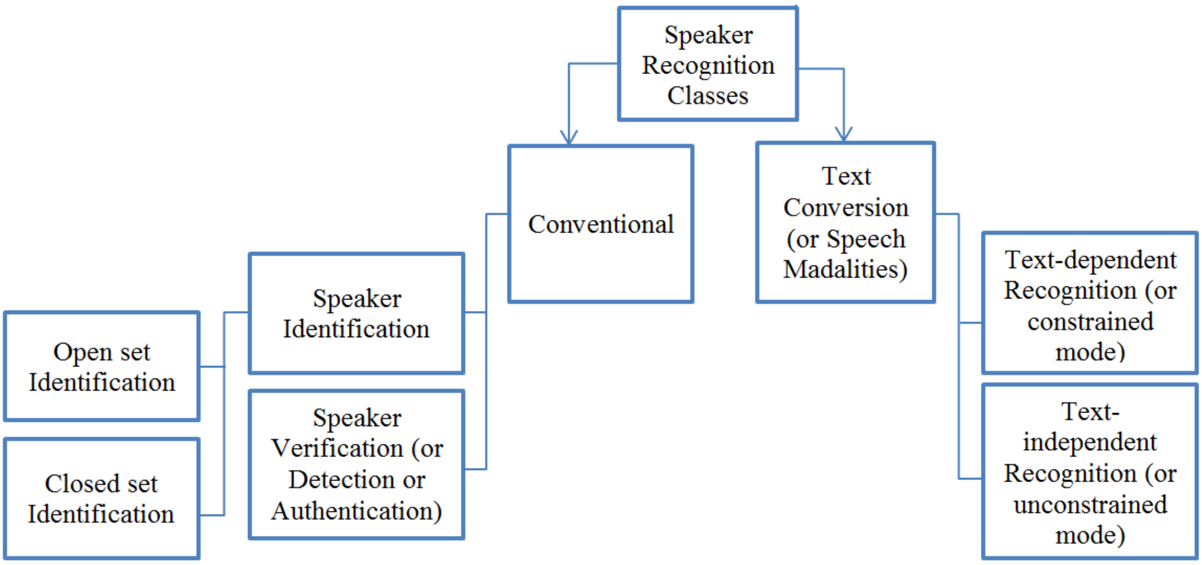
Speaker Recognition Classes
Conventional
Text Conversion (or Speech Madalities)
Speaker Identification
Speaker Verification (or Detection or Authentication)
Open set Identification
Closed set Identification
Text-dependent Recognition (or constrained mode)
Text-independent Recognition (or unconstrained mode)

**Figure 2**: Speaker recognition classes

SR involves two steps: training (or enrollment) and testing (or recognition). In training phase, an input speech utterance is utilized to train a speaker's model. Then, in testing phase, speaker's models collected are used to identify or verify the speaker. SR is categorized into three main methods: artificial neural network and deep learning method, probability model method and, template matching method. For features extraction, the Mel Frequency Cepstral Coefficients (MFCC) is most commonly used. The paper is organized as follows. In Section 2, we describe the state of the art in speaker recognition in general as well as the state of the art in speaker recognition using clustering methods in recent years. Finally in Section 3 the work will be concluded.

## II. State of the Art in Speaker Recognition

Since the early 70s approximately, old works in the state of the art in speaker recognition methods have taken place, and still progressing until now, trying to reach a high accuracy of recognition in the field. Among the contributions related to these works in the past, Fant *et al* (Fant *et al*., 1973) modeled the temporal organization of speech by considering speech as the result of a sequence of phonetic command. Cole *et al* (Cole *et al*., 1983) used several features of human speech and presented them into a system called "FEATURE". Atal *et al* (Atal *et al*., 1983) suggested a temporal decomposition for speech coding. Later, Deleglise *et al* (Deleglise *et al*., 1988) realized localization by implementing the same technique. (Bimbot *et al*., 1991) presented another technique of temporal decomposition by modeling transitions between a succession of acoustic targets, which represent a speech utterance, using interpolations. Bimbot *et al* (Bimbot *et al*., 1992) proposed AR-vector modeling for speech spectral evolution.

Magring-Chagnolleau *et al* (Magring-Chagnolleau *et al*., 1996) explored the performances of AR-vector models for speaker recognition. Campbell *et al* (Campbell *et al*., 1997) described a tutorial of automatic speaker-recognition systems. Trying to refine decisions in speaker identification, Besacier *et al* (Besacier *et al*., 1998) used a hard threshold approach. McLaughlin *et al* (McLaughlin *et al*., 2002) developed a novel Speaker Detection and Tracking (SDT) method combines the traditional training and testing phases. A summary of the SuperSID project to exploit high-level information for high-accuracy speaker recognition was offered by Reynolds *et al* (Reynolds *et al*., 2003). The commonly used temporal and spectral analysis techniques of feature extraction were discussed by (Kesarkar *et al*., 2003) for speech recognition. Lee *et al* (Lee *et al*., 2003) optimized the parameters of the mel-cepstrum transformation in order to minimize the loss of information during the feature extraction stage in speech recognition. State-of-the-art in speaker recognition was provided by Faundez-Zanuy *et al* (Faundez-Zanuy *et al*., 2005). In order to target a specific speaker, Bonastre *et al* (Bonastre *et al*., 2006) explored the effect of a transfer function-based voice transformation on automatic speaker recognition system performance, by emphasizing the enhancing of impostor acceptance rate. Sturim *et al* (Sturim *et al*., 2007a) gave an overview of state-of-the-art automatic speaker recognition systems. To compensate for the effects of auxiliary microphones on the speech signal, Sturim *et al* (Sturim *et al*., 2007b) described two techniques. Through Latent Factor Analysis (LFA) and Nuisance Attribute Projection (NAP), the first method mitigates session effects. With noise reduction techniques, the second approach operates directly on the recorded signal. A new algorithm for the classification of harmonic (short-term periodic) and non-harmonic segments in speech signals was provided by Kavanagh *et al* (Kavanagh *et al*., 2007).

To enhance the SR systems, several techniques have been proposed such as feature extraction, wavelet decomposition etc… For the purpose of improving efficiency of SR, Tiwari *et al* (Tiwari *et al*., 2010) made some changes on the classical Mel Frequency Cepstrum Coefficient (MFCC) technique in order to extract features and to design a text-dependent speaker identification system. Based on a Maximum Likelihood Linear Regression (MLLR) matrix, Sarkar *et al* (Sarkar *et al*., 2010) developed a fast algorithm which computes the likelihood from a model of speaker to rapidly identify the best N speakers. Because of the non-stationary behavior of speech waves, and regarding the good performance of wavelet based approach, Rajakumar *et al* (Rajakumar *et al*., 2010) emphasized on wavelet decomposition to yield a better speech signals time-frequency characteristics, aiming to improve speech models. For the goal of creating a large binary vector and improving the classical processes using a discrete or binary view, Bonastre *et al* (Bonastre *et al*., 2011) suggested a novel method to structure the acoustic space in regions related on a Universal Background Model (UBM). To emphasize speaker specific information, a set of Gaussian models (called specificities) populated each region. Anusuya *et al* (Anusuya *et al*., 2011) discussed the accuracy enhancement of signal recognition applying the pre-processing using wavelet techniques to the conventional methods. To achieve the user recognition, Lin *et al* (Lin *et al*., 2012) designed a fusion recognition system, based on the idea of confidence indices that combines face and speaker classifiers. Das *et al* (Das *et al*., 2014) developed a text-dependent speaker recognition system using MFCC for features extraction and Vector Quantization (Linde Buzo Gray) VQ (LBG) for design of codebook from extracted features. Euclidean distance was used for the computation of VQ distortion derived between the utterances of unknown speaker to codebooks of known speaker. McLaren *et al* (McLaren *et al*., 2014) investigated a new algorithm for speaker recognition through the two-dimensional Discrete Cosine Transform (2D-DCT) instead of MFCC to contextualize features. Gender recognition of speech with an accuracy analysis was the aim of Maka *et al* (Maka *et al*., 2014) study. While a combination of Linear Prediction Coding (LPC) with Mel Frequency Cepstral Coefficients (MFCC) algorithms was applied by Bansod *et al* (Bansod *et al*., 2014) in order to extract speech features for identification task.

In the next section, we will describe basically the main contributions of speaker recognition based clustering algorithms since the early 90s until 2015.

## III. State of the Art in Speaker Recognition Using Clustering Methods

State of the art in speaker recognition using clustering methods has greatly progressed in the latest years. In this work, we include the very popular clustering techniques. Their state of the art is described as follows:

### III.1. Speaker Recognition Using Fuzzy C-Means (FCM)

For fuzzy clustering, and for Automatic Speaker Recognition (ASR), Jang *et al* (Jang *et al*., 1997) presented the use of neuro-fuzzy and soft computing techniques. Hossan *et al* (Hossan *et al*., 2011) developed a new automatic speaker recognition approach for speaker modeling called Fuzzy Vector Quantization (FVQ) by integrating Vector Quantization (VQ) based on the fuzzy c-means clustering algorithm. In noisy conditions, the performance of MFCC+ Gaussian Mixture Model (GMM) and MFCC+GMM+UBM decreases for speaker identification when the population size is increased. To alleviate this problem, Hu *et al* (Hu *et al*., 2013) suggested a fuzzy-clustering-based decision tree approach.

### III.2. Speaker recognition using Gaussian Mixture Model (GMM)

Concerning the GMMs, and due to the limitations of GMM/UBM (Gaussian Mixture Model/ Universal Background Model) based systems in ergonomic constraints and a limited amount of computing resources, Larcher *et al* (Larcher *et al*., 2008) offered a method to address these limitations. For the purpose of getting additional discriminant features, Anguera *et al* (Anguera *et al*., 2012) proposed the combination of GMM posteriorgrams front-end with K-means clustering method. Kaminski *et al* (Kaminski *et al*., 2013) offered an automatic speaker recognition system which utilizes GMM in the classification process and a unique feature vector Voice Print (VP) in defining the voice. For robust automatic speech recognition, Liu *et al* (Liu *et al*., 2014) described a new algorithm by a combination of Deep Neural Network (DNN) and GMM with unsupervised speaker adaptation using the Temporally Varying Weight Regression framework, to take profit of the strong adaptability of the GMMs and the high performance of the DNNs. Yunqi *et al* (Yunqi *et al*., 2014) presented and discussed the application to speaker recognition of a novel method called adaptive Gaussian mixture model (AGMM) based on the actual distributions of characteristic parameter.

### III.3. Speaker recognition using Hierarchical Clustering

Dealing with hierarchical clustering, a hierarchical mixture clustering technique was suggested by Saeidi *et al* (Saeidi *et al*., 2008). Garcia-Romero *et al* (Garcia-Romero *et al*., 2014a) introduced a structure based i-vector speaker recognition systems for unsupervised domain adaptation of Probabilistic Linear Discriminant Analysis (PLDA), and by the way two versions of agglomerative hierarchical clustering that use the PLDA system were explored.

### III.4. Speaker recognition using Principal Component Analysis (PCA)

Concerning methods of speaker recognition using data reduction by Principal Component Analysis (PCA), and in order to extract and identify speech features, Nie *et al* (Nie *et al*., 2004) combined principle component analysis with an artificial neural network approach. A new technique based on PCA classifier and Kernel Fisher Discriminant (KFD) classifier was offered by Li *et al* (Li *et al*., 2008). This algorithm represents a novel hierarchical speaker verification. On the other hand, Zhao *et al* (Zhao *et al*., 2009) developed a novel technique based on PCA and vector quantization so as to solve the slow recognition response speed in large population. Based on a combination of Discrete Stationary Wavelet Transform (DSWT) with Principal Component Analysis (PCA) techniques, Jayakurnar *et al* (Jayakurnar *et al*., 2009) developed a robust algorithm for speaker identification. Zhou *et al* (Zhou *et al*., 2010) introduced a method to reduce feature dimension which includes the use of Canonical Correlation Analysis (CCA) in order to fuse the Linear Predictive Coding (LPC) and MFCC features and the use of PCA in order to decrease the effective features dimension and eliminate features redundancy. Xiao-chun *et al* (Xiao-chun *et al*., 2012) suggested a text-independent speaker identification to suppress the phonetic information by using a subspace method. The subspaces will be constructed by Probabilistic Principle Component Analysis (PPCA). To extract mixed characteristic parameters, Jing *et al* (Jing *et al*., 2014) proposed a novel method of speaker recognition using PCA, selecting the combination of Linear Prediction Cepstrum Coefficient (LPCC) and MFCC, and the first-order differential parameter as the characteristic parameter.

### III.5. Speaker recognition using Support Vector Machines (SVMs)

Using Support Vector Machines (SVM) technique, Louradour *et al* (Louradour *et al*., 2008), provided a comparative study of three State-of-the- art SVM speaker verification systems based on sequence kernels: the Generalized Linear Discriminant Sequence (GLDS) kernel, the GMM-supervectors sequence kernel and the Feature Space Normalized Sequence (FSNS) kernel, then they offered a comparison of these three SVM systems to the conventional generative Universal Background Model (UBM)-GMM. Gonzalez *et al* (Gonzalez *et al*., 2013) presented a new procedure to perform speaker likability classification by identifying a small set of features that will be incorporated in a linear support vector machine (SVM).

### III.6. Speaker recognition using Neural Networks (NN)

Employing Neural Networks, Tebelskis et al (Tebelskis et al., 1995) studied the advantage of applying neural networks on continuous speech recognition system, a large vocabulary and on speaker independent. A rapid speaker adaptation algorithm for automatic speech recognition systems based on artificial neural networks (ANNs) for Hidden Markov Models (HMMs) state probability estimation was developed by Dupont et al (Dupont et al., 2000). A comparative study of neural network classifiers with PCA and lip features for speaker identification task was given by Mehra *et al* (Mehra *et al.*, 2010). Graves et al (Graves et al., 2013) introduced and assessed the potential of deep Long Short-term Memory Recurrent Neural Networks (RNNs) for speech recognition and presented an improvement of end-to-end learning method which trains jointly as acoustic and linguistic models two distinct Recurrent Neural Networks (RNNs). For text independent speaker recognition, Ahmad et al (Ahmad et al., 2015) presented a unique approach using the combination of the MFCC with its delta derivatives (DMFCC and DDMFCC) calculated using mel spaced Gaussian filter banks and described speaker modeling using Probabilistic Neural Network (PNN) in order to reach lower operational times. For feature dimensionality reduction, PCA is applied before the enrollment and recognition phases.

From our investigation in dealing with speaker recognition process, we notice that recent studies (from 2014 to 2015) were focusing on Deep Neural Networks **(**DNNs). Among these works, we cite that Lei *et al* (Lei *et al.*, 2014) suggested a new algorithm for speaker recognition based on Automatic Speech Recognition (ASR)-DNN system where the statistics for the i-vector model are extracted by DNNs trained for ASR. Because of the exceptional and unique results reached by DNNs, Romero *et al* (Garcia-Romero *et al.*, 2014b) inspected the use of DNNs to collect Sufficient Statistics (SS) for the unsupervised domain adaptation task of the Domain Adaptation Challenge (DAC) instead of GMMs in the traditional i-vector speaker recognition frameworks. McLaren *et al* (McLaren *et al.*, 2015) applied DNN-based Speaker Identification (SID) on microphone speech. Using the 2013 Domain Adaptation Challenge Speaker Recognition (DAC13) and the NIST 2011 Language Recognition Evaluation (LRE11) benchmarks, Richardson *et al* (Richardson *et al.*, 2015) applied single DNN for both Speaker Recognition (SR) and Language Recognition (LR).

## IV. Conclusions and Future work

In this study, the principal contributions of the state of the art in speaker recognition based on different techniques (MFCC, wavelet transform, 2D-DCT, LPC, …) and based on different clustering techniques (FCM, GMM, PCA, …) were mainly described since the beginning of the 70s to 2015 to have a general idea on the development of speaker recognition methods 45 years ago.

From the literature review in speaker recognition based on different techniques, we notice that in the early 70s, speaker recognition research was principally focusing on temporal decomposition technique. This method was modified and improved during almost 20 years. During the Nineties, the exploration of the performances of AR-vector models for speaker recognition was carried out. From 2000, several techniques were performed. To extract speech features, numerous modifications were made on the classical MFCC technique, for instance, the mel-cepstrum transformation was optimized in order to minimize the loss of information during the feature extraction, a combination of LPC with MFCC algorithms was implemented and 2D-DCT was applied instead of MFCC to contextualize features. Aiming to improve speech models, wavelet decomposition was implemented. Methods based on MLLR, UBM were also suggested to enhance speaker recognition. On the other side, in the 90s, it was the beginning of employing the Neural Networks and the HMMs, while in the 2000s, novel methods were proposed in order to increase the speaker recognition accuracy. Among these new ideas, we can cite: the incorporation of Vector Quantization based on the fuzzy c-means clustering, the combination of GMM with K-means technique, the development of the HMM based on the PPCA, the suggestion of a hierarchical mixture clustering algorithm, the development of a new procedure called DCLM, the combination of PCA classifier with KFD classifier for a hierarchical speaker verification, the combination of PCA and vector quantization and the combination of DSWT with PCA.

Nowadays, the majority of the employed techniques are mainly based on Neural Networks. Among the proposed contributions in this context, we have the combination of DNNs with GMMs. Also, we notice the implementation of the RNNs and the PNN in the recent works and the application of DNNs was the most commonly used because of its unprecedented, exceptional and unexpected results.

As a further contribution, the scientific community could attempt to improve the speaker recognition performances by incorporating Visual Analytics (VA) techniques and fusion of classifiers so as to exploit the speaker characteristics in a better way.

## References

Ahmad, K. S., Thosar, A. S., Nirmal, J. H., Pande, V. S., 2015. A Unique Approach in Text Independent Speaker Recognition using MFCC Feature Sets and Probabilistic Neural Network. Eighth International Conference on Advances in Pattern Recognition (ICAPR), 2015, pp. 1 - 6.

Anguera, X., 2012. Speaker Independent Discriminant Feature Extraction for Acoustic Pattern-Matching. ICASSP 2012, pp. 485– 488.

Anusuya, M. A., Katti, S. K., 2011. Comparison of Different Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition. International Journal of Computer Applications (0975 – 8887), Vol 26– no.4, pp. 19 - 24. http://www.ijcaonline.org/volume26/number4/pxc3874242.pdf.

Atal, B. S., 1983. Efficient coding of LPC parameters by temporal decomposition. IEEE International Conference on Acoustics, Speech, and Signal Procrssing (ICASSP'83), vol. 8, pp. 81- 84.

Bansod, N. S., Dadhade, S. B., Kawathekar, S. S., Kale, K. V., 2014. Speaker Recognition using Marathi (Varhadi) Language. 2014 International Conference on Intelligent Computing Applications, pp. 421- 425.

Besacier, L., Bonastre, J. F., 1998. Time and frequency pruning for speaker identification. Fourteenth International Conference on Pattern Recognition, 1998. Proceedings, Vol 2, pp. 1619- 1621.

Bimbot, F., Atal, B. S., 1991. An evaluation of temporal decomposition. EUROSPEECH.

Bimbot, F., Mathan, L., De Lima, A., Chollet, G., 1992. Standard and Target Driven AR-Vector Models for Speech Analysis and Speaker Recognition. IEEE International Conference on Acoustics, Speech, and Signal Procrssing, 1992, ICASSP-92, Vol 2, pp. 5- 8.

Bonastre, J. F., Bousquet, P. M., Matrouf, D., Anguera, X., 2011. Discriminant Binary Data Representation For Speaker Recognition. IEEE, ICASSP 2011, pp. 5284- 5287.

Bonastre, J. F., Matrouf, D., Fredouille, C., 2006. Transfer Function-Based Voice Transformation For Speaker Recognition. IEEE Odyssey 2006: The Speaker and Language Recognition Workshop, pp. 1- 6.

Campbell, JR. J. P., 1997. Speaker Recognition: A Tutorial. Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462.

Cole, R. A., Stern, R. M., Phillips, M. S., 1983. Feature-Based Speaker-Independent Recognition of Isolated English Letters. IEEE, ICASSP 83, pp. 731- 733.

Das, A., Jena, M. R., Barik, K. K., 2014. Mel-Frequency Cepstral Coefficient (MFCC) - a Novel Method for Speaker Recognition. Digital Technologies, 2014, Vol. 1, no. 1, pp. 1- 3.

Deleglise, P., Bimbot, F., Montacie, C., Chollet, G., 1988. Temporal Decomposition and Acoustic-Phonetic Decoding for the Automatic Recognition of Continuous Speech. 9th International Conference on Pattern Recognition, vol.2, pp. 839- 841.

Dupont, S., Cheboub, L., 2000. Fast speaker adaptation of artificial neural networks for automatic speech recognition. Proceedings. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP '00, Vol 3, pp. 1795- 1798.

Fant, G., 1973. Speech sounds and features. MIT Press.

Faundez-Zanuy, M., Monte-Moreno, E., 2005. State-of-the-Art in Speaker Recognition. IEEE Aerospace and Electronic Systems Magazine, Vol 20, pp. 7 - 12.

Garcia-Romero, D., McCree, A., Shum, S., Brümmer, N., Vaquero, C., 2014a. Unsupervised domain adaptation for I-Vector speaker recognition. Odyssey 2014: The Speaker and Language Recognition Workshop, pp. 260- 264. https://groups.csail.mit.edu/sls/publications/2014/garcia-romero_odyssey2014.pdf.

Garcia-Romero, D., Zhang, X., McCree, A., Povey, D., 2014b. Improving Speaker Recognition Performance in the Domain Adaptation Challenge Using Deep Neural Networks. 2014 IEEE Spoken Language Technology Workshop (SLT), pp. 378 - 383.

Gonzalez, S., Anguera, X., 2013. Perceptually Inspired Features for Speaker Likability Classification. ICASSP 2013, pp. 8490– 8494.

Graves, A., Mohamed, A., Hinton, G., 2013. Speech Recognition With Deep Recurrent Neural Networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6645-6649. http://www.cs.toronto.edu/~fritz/absps/RNN13.pdf.

Hossan, M. A., 2011. Automatic Speaker Recognition Dynamic Feature Identification and Classification using Distributed Discrete Cosine Transform Based Mel Frequency Cepstral Coefficients and Fuzzy Vector Quantization. Master Thesis of Engineering, Electrical and Computer Engineering, College of Science, Engineering and Heath, RMIT University.

Hu, Y., Wu, D., Nucci, A., 2013. Fuzzy-Clustering-Based Decision Tree Approach for Large Population Speaker Identification. IEEE Transactions on Audio, Speech, and Language Processing, VOL. 21, NO. 4, pp. 762- 774.

Jang, J. S. R., Chen, J. J., 1997. Neuro-Fuzzy and Soft Computing for Speaker Recognition. FUZZ-IEEE’97, pp. 663- 668.

Jayakurnar, A., Vimal Krishnan, V. R., Babu Anto, P., 2009. Text dependent speaker recognition using discrete stationary wavelet transform and PCA. 2009 International Conference on the Current Trends in Information Technology (CTIT), pp. 1– 4.

Jing, X., Ma, J., Zhao, J., Yang, H., 2014. Speaker Recognition Based on Principal Component Analysis of LPCC and MFCC. 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 403- 408.

Kaminski, K., Majda, E., Dobrowolski, A. P., 2013. Automatic speaker recognition using a unique personal feature vector and Gaussian Mixture Models. The Institute Of Electrical And Electronics Engineers INC, Signal Processing, Algorithms, Architectures, Arrangements, And Applications, SPA 2013, pp. 220- 225.

Kavanagh, D. F., Boland, F., 2007. A Non-Linear Operator Based Method for Harmonic Feature Extraction From Speech Signals. 2007 IEEE International Conference on Signal Processing and Communications (ICSPC 2007), pp. 217- 220.

Kesarkar, M. P., Rao, P., 2003. Feature Extraction for Speech Recogniton. M.Tech. Credit Seminar Report, Electronic Systems Group, pp. 1- 11. https://www.ee.iitb.ac.in/~esgroup/es_mtech03_sem/sem03_paper_03307003.pdf.

Larcher, A., Bonastre, J. F., Mason, J. S. D., 2008. Short Utterance-based Video Aided Speaker Recognition. IEEE, MMSP 2008, pp. 897- 901.

Lee, C., Hyun, D., Choi, E., Go, J., Lee, C., 2003. Optimizing Feature Extraction for Speech Recognition. IEEE Transactions On Speech and Audio Processing, Vol. 11, no. 1, pp. 80- 87.

Lei, Y., Scheffer, N., Ferrer, L., McLaren, M., 2014. A Novel Scheme for Speaker Recognition Using a Phonetically-Aware Deep Neural Network. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1695 - 1699.

Li, M., Xing, Y., Luo, R., 2008. Hierarchical Speaker Verification Based on PCA and Kernel Fisher Discriminant. Fourth International Conference on Natural Computation, pp. 152- 156.

Lin, C. Y., Song, K. T., Chen, Y. W., Chien, S. C., Chen, S. H., Chiang, C. Y., Yang, J. H., Wu, Y. C., Liu, T. J., 2012. User Identification Design by Fusion of Face Recognition and Speaker Recognition. 2012 12th International Conference on Control, Automation and Systems, pp. 1480- 1485.

Liu, S., Sim, K. C., 2014. On Combining DNN and GMM With Unsupervised Speaker Adaptation for Robust Automatic Speech Recognition. 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), pp. 195- 199.

Louradour, J., Daoudi, K., 2008. State-of-the-art sequence kernels for SVM speaker verification. IEEE Workshop on Machine Learning for Signal Processing, 2008. MLSP 2008, pp. 498 - 503.

Magring-Chagnolleau, I., Wilke, J., Bimbot, F., 1996. A Further Investigation on AR-Vector Models for Text-Independent Speaker Identification. Conference Proceedings., 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-96, vol. 1, pp. 101- 104.

Maka, T., Dziurzanski, P., 2014. An Analysis of the Influence of Acoustical Adverse Conditions on Speaker Gender Identification. 2014 XXII Annual Pacific Voice Conference (PVC), pp. 1 - 4.

Matrouf, D., Bonastre, J. F., 2006. Accurate Log-Likelihood Ratio Estimation By Using Test Statistical Model For Speaker Verification. IEEE Odyssey 2006: The Speaker and Language Recognition Workshop, pp. 1- 5.

McLaren, M., Lei, Y., Ferrer, L., 2015. Advances in Deep Neural Network Approaches to Speaker Recognition. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4814 - 4818.

McLaren, M., Scheffer, N., Ferrer, L., Lei, Y., 2014. Effective Use of DCTS For Contextualizing Features For Speaker Recognition. 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), pp. 4027- 4031.

McLaughlin, J., Reynolds, D. A., 2002. Speaker Detection and Tracking for Telephone Transactions. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol 1, pp. I-129 - I-132.

Mehra, A., Kumawat, M., Ranjan, R., Pandey, B., Ranjan, S., Shukla, A., Tiwari, R., 2010. Expert System for Speaker Identification using Lip features with PCA. 2010 2nd International Workshop on Intelligent Systems and Applications (ISA), pp. 1– 4.

Nie, K., Zeng, F. G., 2004. Using Neural Network and Principal Component Analysis to Study Vowel Recognition with Temporal Envelope Cues. Proceedings of the 26th Annual International Conference of the IEEE EMBS, pp. 4592- 4595.

Rajakumar, P. S., Ravi, S., Suresh, R. M., 2010. Speech Enhancement Models Suited For Speech Recognition Using Composite Source And Wavelet Decomposition Model. 2010 International Conference on Signal and Image Processing, pp. 511- 514.

Reynolds, D., Andrews, W., Campbell, J., Navratil, J., Peskin, B., Adami, A., Jin, Q., Klusacek, D., Abramson, J., Mihaescu, R., Godfrey, J., Jones, D., Xiang, B., 2003. The SuperSID Project: Exploiting High-level Information for High-accuracy Speaker Recognition. ICASSP 2003, pp. 784- 787.

Richardson, F., Reynolds, D., Dehak, N., 2015. Deep Neural Network Approaches to Speaker and Language Recognition. IEEE Signal Processing Letters, Vol: 22, pp. 1671 - 1675.

Saeidi, R., Sadegh Mohammadi, H. R., Ganchev, T., Rodman, R. D., 2008. Hierarchical mixture clustering and its application to GMM based text independent speaker identification. 2008 Internatioal Symposium on Telecommunications, pp. 770- 773.

Sarkar, A. K., Rath, S. P., Umesh, S., 2010. Fast Approach to Speaker Identification for Large Population using MLLR and Sufficient Statistics. 2010 National Conference on Communications (NCC), pp. 1 - 5.

Sturim, D. E., Campbell, W. M., Reynolds, D. A., 2007a. Classification Methods for Speaker Recognition. Springer-Verlag Berlin Heidelberg, C. Müller (Ed.): Speaker Classification I, LNAI 4343, pp. 278–297.

Sturim, D. E., Campbell, W. M., Reynolds, D. A., Dunn, R. B., Quatieri, T. F., 2007b. Robust Speaker Recognition with Cross-Channel Data: MIT-LL Results on the 2006 NIST SRE Auxiliary Microphone Task. ICASSP 2007, pp. 49- 52.

Tebelskis, J., 1995. Speech Recognition using Neural Networks. PhD Thesis of Philosophy in Computer Science, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Tiwari, V., 2010. MFCC and its applications in speaker recognition. International Journal on Emerging Technologies, pp. 19- 22. http://researchtrend.net/ijet/4_Vibha.pdf.

Xiao-chun, L., Jun-xun, Y., 2012. A Text-independent Speaker recognition System Based on Probabilistic Principle Component Analysis. 2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization, pp. 255- 260.

Yunqi, W., Yibiao, Y., 2014. Accurate Speaker Recognition Based On Adaptive Gaussian Mixture Model. ICSP2014 Proceedings, pp. 527- 531.

Zhao, Z. D., Zhang, J., Tian, J. F., Lou, Y. Y., 2009. An effective identification method for speaker recognition based on PCA and double VQ. Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, pp. 1686– 1689.

Zhou, Y., Zhang, X., Wang, J., Gong, Y., 2010. Research on speaker feature dimension reduction based on CCA and PCA. 2010 International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1– 4.