

## A novel automatic voice recognition system based on text-independent in a noisy environment

Motaz Hamza, Touraj Khodadadi, Sellappan Palaniappan

Department of Information Technology, Malaysia University of Science and Technology, Malaysia

---

### Article Info

#### Article history:

Received Jul 15, 2019

Revised Jan 8, 2020

Accepted Jan 29, 2020

---

#### Keywords:

Automatic voice recognition (AVR)

Gaussian mixture model (GMM)

Mel frequency cepstral coefficients (MFCC's)

Microsoft research (MSR) identity toolbox

Universal background model (UBM)

---

### ABSTRACT

Automatic voice recognition system aims to limit fraudulent access to sensitive areas as labs. Our primary objective of this paper is to increase the accuracy of the voice recognition in noisy environment of the Microsoft Research (MSR) identity toolbox. The proposed system enabled the user to speak into the microphone then it will match unknown voice with other human voices existing in the database using a statistical model, in order to grant or deny access to the system. The voice recognition was done in two steps: training and testing. During the training a Universal Background Model as well as a Gaussian Mixtures Model: GMM-UBM models are calculated based on different sentences pronounced by the human voice (s) used to record the training data. Then the testing of voice signal in noisy environment calculated the Log-Likelihood Ratio of the GMM-UBM models in order to classify user's voice. However, before testing noise and de-noise methods were applied, we investigated different MFCC features of the voice to determine the best feature possible as well as noise filter algorithm that subsequently improved the performance of the automatic voice recognition system.

Copyright © 2020 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Touraj Khodadadi,

Department of Information Technology,

Malaysia University of Science and Technology,

12, Jalan PJU 5/5, Kota Damansara, 47810 Petaling Jaya, Selangor, Malaysia.

Email: touraj@must.edu.my

---

## 1. INTRODUCTION

The process of voice recognition starts with recording the user's voice through the microphone. In a default mode, noise from the environment is automatically recorded together with the voice. Subsequently, the de-noising method is applied to the recorded voice. There are a number of features of noise extraction that can be applied to the de-noise speech signal, usually, (MFCC's) would be applied. The extracted features characterise the frequency content of the voice, the shape of the vocal tract, and the intonation, or prosody. For each voice frame, a vector of 20 to 30 characteristics, termed as the "Cepstral" coefficients are calculated [1]. Finally, the classification of the users' voices is done using the Log-Likelihood Ratio (LLR) of the UBM-GMM model. However, before being used in the classification of the users' voices, the UBM-GMM model is created during the training process using the training data which contains speech signals from known users [2]. The recognition of a speaker through voice performance is subjected to signal quality, and it also depends on the variability of the speaker's voice over time as in the case of illness (colds), emotional states (anxiety, joy, etc), age, voice acquisition conditions (such as noise and reverberation, the quality of equipment such as the microphone) [3].

MSR identity toolbox is a very accurate tool for voice recognition when tested in a quiet (no noise) environment [4]. However, the performance of the toolbox is greatly reduced when tested in a noisy environment, according to the study by [5], which investigates the relations between the accuracy of speaker

recognition and adverse acoustic conditions. The researcher established that the accuracy of the MSR identity toolbox is significantly reduced when used in a noisy environment, the author documented the implication of noise and reverberation in the research. In particular, a regression formula has been established to predict the recognition accuracy of a typical speech recognition system. On the other hand, Dirceu presents an evaluation of the Microsoft Research Identity Toolbox under conditions reflecting those of a real-world condition which means noisy environment, and he established that the accuracy of the MSR identity toolbox drops when tested in a noisy environment [6-9]. However, both researchers did not provide solutions to this problem. Hence, the research is conducted to improve the accuracy of (MSR) identity toolbox in a noisy environment. The research attempts to address the issues in the domain of voice recognition in a noisy environment, by:

- Using several noise removal (de-noising) methods to improve the MSR identity toolbox;
- Determining the best MFCC that suits both voice recognition and noise removal, since multiple types of MFCC voice features may be used.

## 2. RESEARCH OBJECTIVES

This research objectives: (a) To improve the Microsoft research team's MSR identity toolbox and making it robust in a noisy environment. A method based on noise removal filtering and the choice of the best MFCC speech signal feature that gives the best accuracy for voice recognition in a noisy environment is proposed. Firstly, we implemented noise removal filters such as Filter bank, least mean squares filter, Wavelet filter, and hearing aid filter in order to improve the accuracy of the MSR identity toolbox in real-life conditions (noisy environment). Subsequently, we investigated several types of the MFCC voice features to show their effects on the accuracy of the voice recognition system, and to determine the best MFCC that suits both UBM-GMM voice recognition model and the noise removal methods for better accuracy. (b) To develop the modules of the proposed voice recognition system using MSR identity toolbox, which implements the GMM-UBM model. Several modules will be implemented in MATLAB, such as Users enrolment, management, users' voice recording, a module for training and testing, a module for noise application and removal, MFCC features extraction module, and a module for wav files conversion in the case of usage of TIMIT voice database. (c) To evaluate the feasibility of the proposed method in improving the voice detection accuracy of the MSR identity toolbox in a noisy environment.

## 3. RELATED WORKS OF VOICE RECOGNITION

The first research work for AVR was done in 1985, it uses: Vector Quantization (VQ) [10-12] from the enlistment data of a given voice, a partitioning of the acoustic space into a finite number of regions and each represents a centroid vector. The set of these vectors forms what are called a quantization dictionary. The proximity measurement between this dictionary and the frames of a test segment, calculated as the average of the minimum distances between each frame and the centroids, allows the comparison of voice segments. This method can be described as deterministic, but also vectorial and non-parametric. Its results depend strongly on the fixed size of the dictionary. Core approaches, such as Support Vector Machine SVM [13-16] apply to frame a nonlinear transformation to a large superspace. This method is to maximize the linear margin between classes. The performance of SVM-based systems now competes with those of the state of the art Gaussian mixture that we will present later. This method is qualified as semi-deterministic (the transformation may or may not derive from theoretical laws), vector and discriminant. Anchor models [17] represent a voice relative to a model space, in the form of a vector of similarities with respect to them. A similarity can be calculated by a comparison score (anchor model VS voice learning data).

This method is strongly dependent on the robustness of the anchoring models. It depends on their ability to synthesize the structural information about the acoustic space by a discrete overlap. Other methods have been presented for about twenty years, as well as variants of the previous methods. They are generative or discriminative "oriented": Dynamic Bayesian Networks (DBN) [18-19], Segment Mixture Models (SMM) [20] combining GMM approaches (which we propose below) and Dynamic Time Warping (DTW) [21] to take into account the temporal structure of the voice signal. The GDW (Gaussian Dynamic Warping) approach of [22], which is a hybrid exploiting the generalization of generative approaches and the structural modeling of DTW. VQ codebook model is developed based on K-means clustering procedure for these perceptual features. In this algorithm there is a mapping from L training vectors into M clusters. Each block is normalized to unit magnitude before giving as input to the model. One model is created for each human voice [23]. In the family of discriminating linear classifiers, [24] introduced the classifier by logistic regression, the latter having the advantage of relying on probabilistic and Bayesian hypotheses. However, its performance does not match those of the SVM in the experiments of [24]. Also noteworthy is

a recent discriminatory linear classifier approach based on SVM, Pairwise SVMs [25], which goes beyond the limits of SVMs based on the supervisors. Modeling based on GMM is part of the more general Hidden Markov Models (HMM) model. Hidden Markov models apply perfectly to text-dependent mode, obtaining excellent results. On the other hand, the use of HMM models in text-independent mode does not improve the performance achieved by simpler models based on GMM [26]. These use the notions of state introduced by the Russian mathematician. Their succession, with transition probabilities, and makes it possible to elaborate a stochastic model of the voice. This approach, initially used for voice recognition, is very effective for voice recognition. According to the literature review paper done by [2] GMM based AVR is the best method because it is excellent for voice pattern recognition problem. In addition [27] developed a GMM based automatic voice verification system development for forensics in Bahasa Indonesia. Furthermore [28, 29] have proved from their research work that GMM is the best for voice recognition, then followed by DTW and finally the less performing algorithm, namely SVM. As shown in Figure 1, the researchers compared GMM, DTW and SVM using the TIMIT voice corpus. They increased the number of voices from 10 to 50.

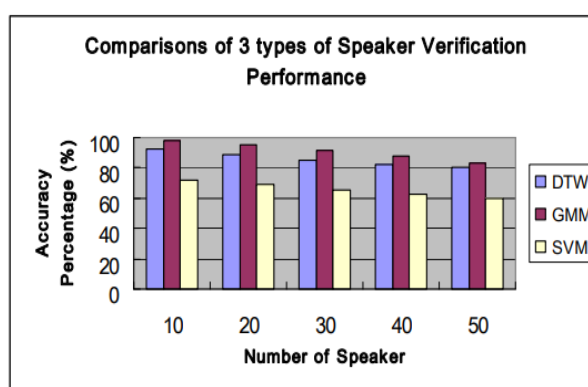


Figure 1. Performance comparisons of GMM, DTW, and SVM

#### 4. RESEARCH METHOD

Voice recognition systems aim to verify and identify unknown voice signal with specific voice signals in the database. Voice recognition systems are classified into two types of services [30]:

- a. Voice verification confirms the identity claim
- b. Voice identification determines of registered voices

Our purposed system is the Voice Verification System (VVS) and it is divided into three key components: feature extraction, voice modeling, and decision.

##### 4.1. Feature extraction of voice signals

Mel Frequency Cepstral Coefficients (MFCCs) is a type of features that describes the characteristics of voice signal, its working is similar to the simulation of the human auditory system that has proved its effectiveness in noise environment; it operates more accurately in frequency domain than the time domain. The MFCC consists of six stages as shown in Figure 2. The first stage in MFCC is the pre-emphasis where increase frequency of voice signal to restore the original voice signal at low levels of noise. The second stage is the frame blocking that divides voice signal into many frames that are used to keep the voice signal for a longer time. The third stage is the hamming window which each frame is multiplied to keep continuing and reduce truncated of voice signal at the first and last points in the frame. The fourth stage is the Fourier transform that used to convert the frame from time domain to frequency domain. The fifth stage is the Mel scale filter that used to obtain high frequency and minimize unwanted frequency. The last stage is the log energy computation, then followed by discrete cosine transform that is applied to return to the time domain [1].

##### 4.2. Voice modeling

The voice modeling step exploits the data provided in the feature extraction step in order to create the representation of an individual who will subsequently be used to authenticate it. The model used is usually a statistical representation of the acquired data. The voice model depends also on the quality of microphones, expected channels of voice signal, amount of voice data duration enrollment and detection.

In this paper, we have chosen the stochastic volatility model such as (GMM-UBM) which is simple, easy to evaluate, and faster to compute classification that is used for human voice verification of the text-independent in noisy environment. As shown in Figures 3 and 4 the UBM model is first created by using Expectation Maximization algorithm (EM) then it is used as feedback to create the GMM model through Maximum A Posteriori algorithm (MAP). Then UBM-GMM models are used in order to recognize the unknown voice during the testing process using the log-likelihood ratio and error equal rate.

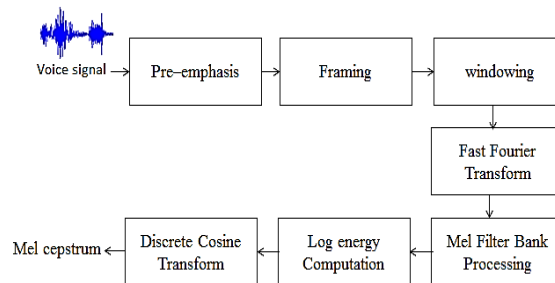


Figure 2. Extraction of parameters with MFCC

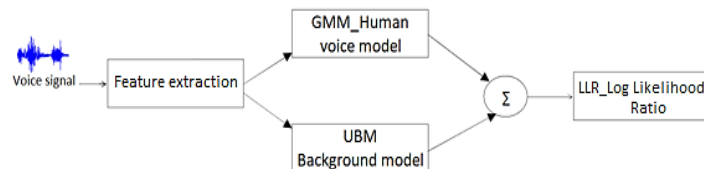


Figure 3. General classifier structure for VVS

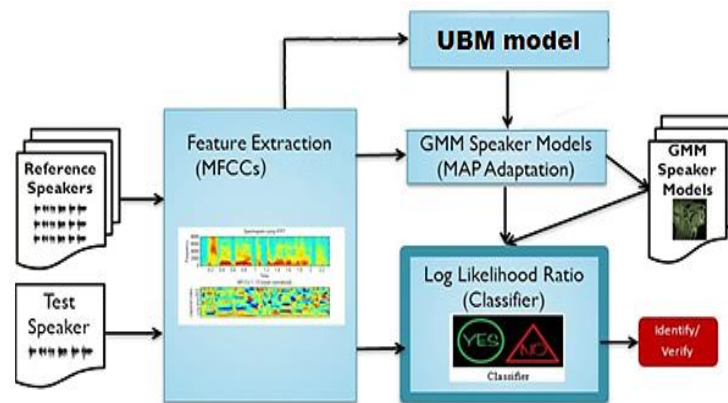


Figure 4. Voice verification system framework

#### 4.3. Decision and performance measures

It consists of checking the adequacy of the vocal message with the acoustic reference of the voice it pretends to be. It is a decision in all or nothing. The performance of voice verification is given in terms of the accuracy rate that defines as a percentage of true voices accepted during the system test.

## 5. RESULTS AND ANALYSIS

### 5.1. Data collection

In the data collection stage, we used the TIMIT database that contains 630 (192 female and 438 male) human voice records with a sampling frequency. Each user has recorded 10 sentences, the voice record duration of each sentence ranged between two to five seconds. We selected 530 human voices randomly as background data and the remaining 100 human voices for users' enrollment data. We have also selected

randomly 9 out of 10 sentences per human voice and the remaining sentence is kept in target human voice data. Then we collected the data for 50 human voices randomly from the users' enrollment data for system testing.

## 5.2. Voice verification system consists of two phases: training and testing

- a. Train phase: After data collection, we have created (MFCCs) for each human voice using Hcopy, it is a tool of HTK to convert the voice signal to different types of MFCCs. In order to calculate MFCCs we have selected the configuration file that we have taken from HTK-MFCC-MATLAB Toolbox to use it with Hcopy tool. Then we trained background data to calculate UBM model by using the algorithm of the MSR identity toolbox, namely Expectation Maximization (EM). Then we applied the Training user enrollment data to calculate GMM model by using the algorithm of the MSR identity toolbox is namely Maximum A Posteriori (MAP).
- b. Test phase: To verify the voice, we added background noise. Then we reduced the noise by using noise filter in order to get an enhanced signal. After that we extracted (MFCCs) feature of the enhanced signal in the noisy environment by using Hcopy tool. Finally, we matched MFCCs feature of enhanced signal with another features of human voices that existed in the users enrollment data by using GMM-UBM trained models with algorithms of MSR identity toolbox, namely Log Likelihood Ratio (LLR) and Error Equal Rate (EER). In this section, all of our experiment results have been done in MATLAB v 8.3. The results have shown the system under perfect conditions (no noise) accuracy was nearly 100% when tested with 50 human voices. Then noise and de-noise methods were applied in order to determine the best configuration of MFCC features that better suits the MSR identity toolbox in order to achieve the best accuracy possible. To do this we changed the MFCC configurations of Hcopy that are configured to automatically convert its wave file input into another types of MFCC. Subsequently, we applied two types of noise namely Background noise and White Gaussian noise at different signal to noise ratios (SNR) (-30 to 60 db), and followed by four different de-noising methods - Filter bank, Least mean squares filter (LMS), Wavelet filter, and Digital hearing aid filter. The implement of the Digital hearing aid filter was the most complicated with many filter types and different steps; therefore, it seems that it is not the most suitable for our system as all its results recognition were zero. From the results listed in the Tables 1 and 2 we notice that Filter bank when choose MFCC\_0\_D\_A, the results of average accuracy at SNR (-30 to 60 db) were: 54.875% for Background noise as well as 14.750% for White Gaussian noise. The results of the other filters are listed in the following tables, then comparison of our proposed system, and with the results of the work of [23] in the clean and noisy environment.

Table 1. Using bankground noise to 50 voices at snr (-30 to 60 db)

Feature type/Filters	No de-noising	Filter bank	Wavelet filter	LMS filter
MFCC_0	42.500%	52.750%	20.750%	11.125%
MFCC_0_D	51.000%	54.250%	38.750%	35.250%
MFCC_0_D_A	51.310%	54.875%	41.500%	46.250%
MFCC_0_D_A_Z	47.875%	48.000%	38.125%	48.750%
MFCC_0_D_Z	46.125%	46.375%	33.250%	48.375%
MFCC_0_Z	42.75%	43.000%	7.1875%	44.25%
MFCC_E	51.625%	52.375%	37.000%	10.250%
MFCC_E_D	53.125%	53.250%	41.375%	36.500%
MFCC_E_D_A	52.750%	52.875%	42.625%	46.875%
MFCC_E_D_A_Z	48.375%	48.625%	40.375%	49.125%
MFCC_E_D_Z	47.625%	48.125%	39.875%	48.625%
MFCC_E_Z	45.250%	45.375%	33.375%	46.500%
MFCC_Z	50.000%	50.375%	32.625%	49.875%

Table 2. Using White Gaussian noise to 50 voices at snr (-30 to 60 db)

Feature type/Filters	No de-noising	Filter bank	Wavelet filter	LMS filter
MFCC_0	5.000%	9.750%	2.870%	3.870%
MFCC_0_D	8.750%	13.000%	3.000%	8.125%
MFCC_0_D_A	10.500%	14.750%	2.620%	11.000%
MFCC_0_D_A_Z	10.250%	10.375%	3.250%	11.500%
MFCC_0_D_Z	10.000%	10.000%	2.750%	11.625%
MFCC_0_Z	9.000%	9.000%	0.375%	9.125%
MFCC_E	11.250%	11.375%	3.125%	3.500%
MFCC_E_D	13.000%	13.125%	5.000%	10.810%
MFCC_E_D_A	14.125%	14.250%	6.750%	13.250%
MFCC_E_D_A_Z	10.250%	10.375%	5.125%	12.000%
MFCC_E_D_Z	10.250%	10.250%	5.250%	12.000%
MFCC_E_Z	9.375%	9.750%	3.375%	10.625%
MFCC_Z	50.000%	50.375%	32.625%	49.875%

The following Figures 5 and 6 show the performance de-noising algorithms under different MFCC using Background noise as well as White Gaussian noise. From Table 3 we notice the result of our proposed system GMM-UBM in clean environment was 100% for 50 voices which is the better result compared to VQ codebook system. The noise deteriorates the information if the voice signals without a noise removal the system accuracy will automatically go down. Therefore, the accuracy of VQ system [23] all will drop in a noisy environment at SNR 15db as shown in the Table 4.

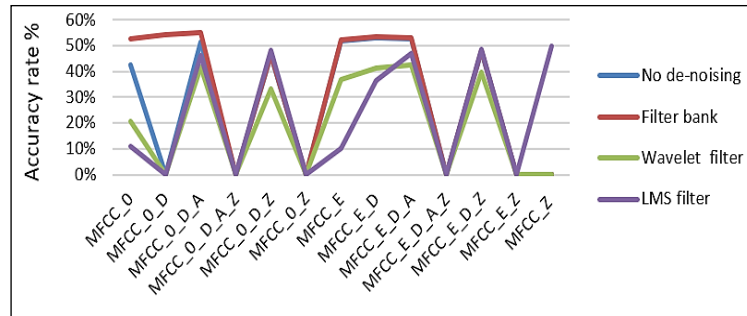


Figure 5. De-noising algorithms performance under different MFCC using background noise

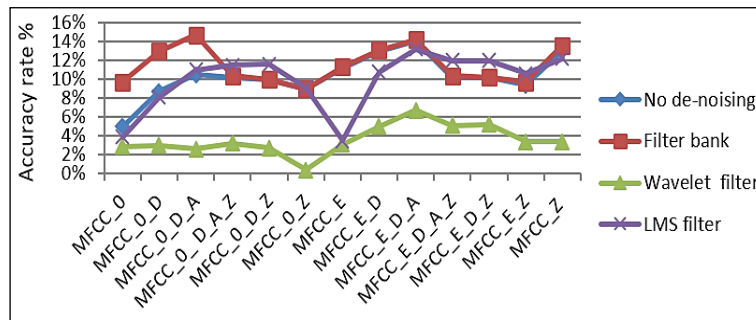


Figure 6. De-noising algorithms performance under different MFCC using White Gaussian noise

Table 3. Comparison of our proposed system with the system of [23] in the clean environment

Reference	Modeling method	Voices database	Number of voices to test	Condition	Accuracy rate in (%)
[23]	VQ codebook	TIMIT	50	Clean	91%
Our proposed system	GMM-UBM	TIMIT	50	Clean	100%

Table 4. Comparison of our proposed system with the system of [23] in the noisy environment at SNR 15db

Reference	Modeling method	Voices database	Number of voices to test	Condition	Accuracy rate in (%)
[23]	VQ codebook	TIMIT	50	Background Noise at 15db	< 91%
Our proposed system	GMM-UBM	TIMIT	50	Background Noise at 15db	100%

From Table 5 we notice the objectives achievement of our proposed system are listed as follows:

- We used MATLAB, implemented our methods and rapid application methodology to achieve our first objective; which is: to design a voice recognition system using the MSR identity toolbox that uses the GMM-UBM models with enhanced modules in MATLAB such as user voice recording.
- To achieve our second objective, we used HTK MFCC configuration to investigate the feature extraction process, and to figure out which configuration of MFCC features that better suits the MSR identity toolbox.

- To achieve the third objective, we implemented our voice recognition system in noisy environment based on MSR identity toolbox, HTK and we implemented our noise removal method (filter bank), in order, to achieve the best accuracy possible compared to VQ codebook system.

Table 5. Results achieved

Reference	Model type	Method	Features extraction	Recognition rate in (%)	Performance	Environments	Noise Filter
[23]	VQ codebook	Iterative Clustering Approach	MF-PLP	91%	Slow	Clean	None
Our proposed system	GMM-UBM	MSR identity toolbox, HTK, and filters noise	MFCC_0_D_A	100%	Fast	Clean & Noisy	Filter bank

## 6. CONCLUSION

In this paper, we presented a robust system of voice recognition by implementing the codes necessary to run the Microsoft research identity toolbox. Most importantly, we have improved the accuracy of the Microsoft research identity toolbox where the result of GMM-UBM was better than VQ codebook in the clean environment by 9%. We found that the result increased the accuracy of the MSR identity toolbox by 3% by applying Filter bank using MFCC\_0\_D\_A for voice recognition system in background noises as well as white Gaussian noise. As a conclusion, our proposed system will contribute to making biometric access through voice a real success in real-world application, as well as playing an important role in the research involving voice recognition technology.

## REFERENCES

- [1] P. M. Chauhan and N. P. Desai, "Mel Frequency Cepstral Coefficients (MFCC) based Speaker Identification in Noisy Environment using Wiener Filter," *presented at Int. Conf. of Green Computing Communication and Electrical Engineering*, 2014.
- [2] J. H. Hansen and T. Hasan, "Speaker Recognition by Machines and Humans: A Tutorial Review," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74-99, 2015.
- [3] S. Memon, *et al.*, "Effect of Clinical Depression on Automatic Speaker Identification," *Presented at The 3rd Int. Conf. On Bioinformatics And Biomedical Engineering*, 2009.
- [4] S. O. Sadjadi, M. Slaney and L. Heck, "MSR Identity Toolbox v1.0: A Matlab Toolbox for Speaker Recognition Research," *Speech Language Process Tech. Committee News*, vol. 1, no 4. 2013.
- [5] K. A. Al-Karawi, A. H. Al-Noori, F. F. Li and T. Ritchings, "Automatic Speaker Recognition System in Adverse Conditions-Implication of Noise and Reverberation on System Performance," *International Journal of Information and Electronics Engineering*, pp. 5-6, 2015.
- [6] D. G. da Silva and , "Evaluation of MSR Identity Toolbox under Conditions Reflecting Those of A Real Forensic Case (forensic\_eval\_01)," *Speech Communication*, pp. 42-49, 2017.
- [7] T. Khodadadi, *et al.*, "Evaluation of Recognition-Based Graphical Password Schemes in Terms of Usability and Security Attributes." *International Journal of Electrical and Computer Engineering*, vol. 6, no. 6, 2016: 2939.
- [8] T. Khodadadi, *et al.*, "Privacy Issues and Protection in Secure Data Outsourcing." *Jurnal Teknologi*, vol. 69, no. 6, 2014.
- [9] F. F. Moghaddam, *et al.*, "VDCI: Variable Data Classification Index to Ensure Data Protection in Cloud Computing Environments," *IEEE Conference on Systems, Process and Control*, 2014.
- [10] F. K. Soong, *et al.*, "Report: A Vector Quantization Approach to Speaker Recognition," *AT&T Technical Journal*, vol. 66, no. 2, pp. 14-26, 1987.
- [11] J. S. Mason, J. Oglesby and L. Xu, "Codebooks to Optimise Speaker Recognition," *1st European Conf. on Speech Communication and Technology*, pp. 1267-1270, 1989.
- [12] P. Mishra, "A Vector Quantization Approach to Speaker Recognition," *Proc. of the Int. Conf. on Innovation & Research in Technology for Sustainable Development*, pp. 152-155, 2012.
- [13] V. N. Vapnik, "An Overview of Statistical Learning Theory," *IEEE Transactions On Neural Networks*, vol. 10, no. 5, pp. 988-999, 1999.
- [14] V. N. Vapnik, "The Nature of Statistical Learning Theory," *Springer Science & Business Media New York*, 2000.
- [15] V. Wan, and W. M. Campbell, "Support Vector Machines for Speaker Verification and Identification," *Neural Networks for Signal Processing X, Proceedings of the 2000 IEEE Signal Processing Society Workshop*, 2000.
- [16] S. Fine, J. Navratil, and R. A. Gopinath, "A Hybrid GMM/SVM Approach to Speaker Identification," *Int. J. of Electrical and Computer Engineering*, vol. 1, no. 4, pp. 721-727, 2007
- [17] T. Merlin, J. F. Bonastre, and C. Fredouille, "Non Directly Acoustic Process for Costless Speaker Recognition and Indexation," *Service Central d'Authentification du CCSD*, 1999.

- [18] G. F. Cooper, and E. Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning," vol. 9, no. 4, pp. 309-347, 1992.
- [19] D. Heckerman, "A Tutorial on Learning with Bayesian Networks," *Learning in Graphical Models*, pp. 301-354, 1999.
- [20] R. P. Stapert and J. S. Mason, "A Segmental Mixture Model for Speaker Recognition," *7th European Conf. on Speech Communication and Technology*, 2001.
- [21] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254-272, 1981.
- [22] J. F. Bonastre, P. Morin and J.C. Junqua, "Gaussian Dynamic Warping (GDW) Method Applied to Text-Dependent Speaker Detection and Verification," *8th European Conf. on Speech Commu. and Technology*, pp. 1-4. 2003.
- [23] A. Revathi, R. Ganapathy, and Y. Venkataramani. "Text Independent Speaker Recognition and Speaker Independent Speech Recognition using Iterative Clustering Approach," *Int. Journal of Computer Science & Information Technology*, vol. 1, no. 2, pp. 30-42, 2009.
- [24] L. Burget, *et al.*, "Discriminatively Trained Probabilistic Linear Discriminant Analysis for Speaker Verification," *Int. Conf. Acoustics, Speech and Signal Processing*, pp. 22-27. 2011.
- [25] S. Cumani, *et al.*, "Pairwise Discriminative Speaker Verification in the I- Vector Space," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 6, pp. 1217-1227, 2013.
- [26] H. Zeinali, H. Sameti, and L. Burget, "Text-Dependent Speaker Verification based on i-vectors, Neural Networks and Hidden Markov Models," *Computer Speech & Language*, vol. 46, pp. 53-71, 2017.
- [27] I. Stefanus, R. J. Sarwono, and M. I. Mandasari, "GMM based Automatic Speaker Verification System Development for Forensics in Bahasa Indonesia," *5th Int. Conf. On Instrumentation, Control And Auto*, 2017.
- [28] Yee, L. M, & Ahmad, A. M. "Comparative Study of Speaker Recognition Methods: DTW, GMM and SVM," 2007.
- [29] M. N. Kakade, and D. B. Salunke. "An Automatic Real Time Speech-Speaker Recognition System: A Real Time Approach." *ICCCE 2019*, pp. 151-158, 2020.
- [30] C. Yu, G. Liu, S. Hahm and J. H. Hansen, "Uncertainty Propagation in Front end Factor Analysis for Noise Robust Speaker Recognition," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014.

## BIOGRAPHIES OF AUTHORS



**Motaz Hamza** is currently doing his M.Sc. in information technology at Malaysia University of Science and Technology (MUST), Malaysia. He received his B.Sc. in Electrical Engineering at Al-Refaq University and High Diploma of Information Technology in Al-Shumoukh Institute in Tripoli, Libya. His main research interests include authentication systems, networks security, cryptography and cybersecurity.



**Dr. Touraj Khodadadi** is a faculty member of the school of science and engineering in Malaysia University of Science and Technology (MUST). He received his Ph.D. and M.Sc. in Information Security in Universiti Teknologi Malaysia (UTM), Malaysia and B.Sc. in Software Engineering from AIU, Iran. Touraj's teaching interests include cyber and network security, authentication systems, cloud computing, big data and machine learning. In addition, he authored and co-authored 30 international journals and conference papers concerning various aspects of computer, information and network security. His main research interests include authentication systems, wireless and wired network security, cryptography, graphical passwords, network authentication, data mining and cloud computing. Apart from teaching and research activities at the university, he has served as an editor and reviewer for several international journals and conferences. He is also members of several review panels for master and doctoral research defense.



**Dr. P. Sellappan** is Professor of Information Technology and Dean of School of Science and Engineering and Provost of the Malaysia University of Science and Technology. He holds a PhD in Interdisciplinary Information Science from the University of Pittsburgh (USA), MSc in Computer Science from the University of London (UK), and BEc in Statistics from the University of Malaya. He has published numerous research papers in reputable international journals and conference papers. He has also authored several IT textbooks for college and university students. His current research interests include Data Mining, Machine Learning, Health Informatics, Blockchain, Cloud Computing, IoT, Cybersecurity and Web Services.