Gender based Voice Authentication Using Gaussian Mixture Model and Mel-Frequency Cepstrum Coefficients

Wahid Rajeh[1]

College of Computers and Information Technology
University of Tabuk, Saudi Arabia

Summary
Biometric is used extensively in the present times in authentication systems with the aim of enhancing security and convertibility. A biometric system used commonly is the voice identification system, which uses the distinct audio features of an individual's voice to identify a person. There are various applications that use this system, for example healthcare services, mobile banking, voice dialing and workforce management. At high noise voice records, there are certain limitations of classical schemes in voice identification applications. It is believed that neural network approach can be used effectively to handle classical schemes problems that emerge in voice identification applications. This study performs a comparison between deep neural network (DNN) and Gaussian Mixture Model (GMM) by determining the accuracy under distinct noise circumstance (i.e., low noise, no noise, heavy noise). The findings depict that there is high accuracy of voice identification system based DNN with distinct noise circumstance, while high noisy voice records are not identified by the GMM model.

Key words:
Biometrics authentication, voice identification, feature matching, Gaussian Mixture Model,Mel-Frequency Cepstrum Coefficients.

1. Introduction

The science of recognizing individuals automatically and uniquely on the basis of one or more of their inherent physical or behavioral attribute is referred to as biometric [1]. Biometrics are of different types, for example fingerprint, voice and iris biometric, and it is possible to group them into two categories, which are physiological biometrics and behavioral biometrics [2]. In physiological biometrics, individuals are recognized on the basis of physiological attributes, like iris scanning and fingerprints, whereas in behavioral biometrics, individuals are identified on the basis of their behavioral attributes, for example voice and signature. Hence, it should be accepted that the biometrics system has turned into a significant tool to identify individuals in different organizations. Though intelligent systems and tools have been developed to identify humans, the risk of being accessed or hacked by someone remains. There is a gap in most systems that hackers can use to exploit and carry out the forgery. The significance of a biometric system is determined by the security and the quality of matching precision it can offer.

Manuscript received January 5, 2022
Manuscript revised January 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.1.70

The characteristic attributes of spoken words or comparing them to one or more voice templates stored in the system is used to achieve voice authentication. Various approaches are used to enhance voice identifications and improve its accuracy. These include Mel-Frequency Cepstrum Coefficients [3], fuzzy logic [4], Gaussian mixture model [5] and lastly using deep neural networks [6]. There are typically two key phases of voice recognition: Training *phase:* The speakers provide their voice to the system in this phase. Hence, the distinct characteristic attributes of speech signal are obtained from the voice. The speaker is then stored in the database with their distinct audio characteristic attributes. *Testing phase:* Unknown users provide their voice sample to the system in this phase. The same approach is then used to obtain the features of the voice. The system then performs a comparison between the audio features of the test sample and the voice samples stored in the database. The user will be provided access if a match is found.

The phonetic features of speech that are distinct for every individual are used by voice recognition. Through this biometric identification technique, users are automatically identified. In addition, the identity of the speaker is linked to the physiological and behavioral attributes of the speaker [2]. It is possible for the same speakers to speak fast, slow or at different speeds, depending on their situation, and often, a vital part is played by the environmental noise in the process of speaker identification. Therefore, the impact of noise on the accuracy of speaker identification systems was determined.

2. The Implemented Model

It was suggested that two speaker identification systems should be analyzed in this project, each of which involving different classifiers, i.e., Deep Neural Network (DNN) and Gaussian Mixture Model (GMM). For this, the accuracy of the two systems would be tested and gauged under distinct noise conditions (heavy noise, low noise, no noise).

Two speaker identification systems were analyzed to verify the model accuracy. For the first system, a technique known as MFSS (Mel Frequency Cepstral Coefficients) was used in combination with Gaussian Mixture Model (GMM), whereas the second system employed MFCC with other features integrated with Deep Neural Network (DNN). The MFCC is used to obtain the characteristics attributes in the speech signal and the GMM and DNN will be trained on these attributes to carry out the matching. MATLAB and Python is used to carry out the experiments across a broad dataset for research on speaker and language recognition that was chosen as the database and referred to as Voxforge. In addition, the dataset was developed and used for analyzing the systems. The accuracy of two speaker identification systems under distinct noise intensity was measured and analyzed. MFCC is used as the feature extractor in the foremost system, along with GMM as a classifier. In the second system, MFCC is used in combination with DNN as a classifier. The comparison was performed under distinct noise conditions to examine the extent to which noise affected the precision of speaker identification systems. The following strategy is adopted when developing the systems: a) select the dataset; b) pre-processing; c) extract features; d) data normalization; e) choose features; and f) identify voice.

3. Related Works

A modeling, risk and trust approach is presented by [8], along with trust approach and evaluation in cognitive identity management by employing biometrics data, digital ID and sensory data. Cognitive checkpoint refers to a semi-automated system that employs AI to process the data sources and to examine trust and risk. There are various elements on which the cognitive system is dependent, for example the perception-cycle, which refers to the information obtained regarding the state of an identified individual, the memory distributed for the whole system, the attention driven by memory to organize the allocation of existing resources, as well as the intelligence acquired through memory, perception and attention. The cognitive approach is used to obtain responses to questions regarding risk of the decision and the trust of the human operator within the system to the decision made by the AI. There are various criteria that form the basis of the risk and trust assessment presented in this working paper, such as information credibility, sensor accuracy, recognition algorithm performance, reliability of sources, etc. It offers various tools for assessing risk and trust, and also for prediction through interference on this network.

A biometric authentication system for improving internet service security was put forward by [9]. The system presented involves a client, the secure voice biometrics server and the application server. The voice and extract features were captured using the client-server, whereas the voices of the user were saved in a database by the biometric server, which were used in the authentication process with the help of a matching algorithm. The Viterbi algorithm provided an EER of 5%. Such biometric is used since passwords and tokens like smart cards are at risk of sharing, theft and loss.

An identification system that employed support vector machine neural networks was presented by [10]. The SVM was trained on the LibriSpeech dataset. The Mel-Frequency Cepstral Coefficients (MFCCs) were used, which are one set of features that depend on frequency representation acquired from audio raw data and the human voice attribute that has frequencies between the range of 85 and 255 Hz. The MFCCs features training the SVM, and the best accuracy attained for the training set was 97%, while it was 95% for the test set. In addition, they also determined that when the time of the audio file was increased from 2 to 4 seconds, the accuracy increased. An authentication framework was presented by [11] for mobile devices, which was known as CORMORANT. The objective of this model was to offer security and privacy of the authentication process. Three biometrics were combined to develop the framework, i.e., face, voice and gait. The risk of unauthorized access decreased in this framework by 97%.

A voice recognition system that employed the Mel-Frequency Cepstral Coefficients (MFCCs) was presented by [12] for the purpose of feature extraction. The Mel-frequency cepstral coefficients refer to the features extracted from the speech data for voice identification. There are widely accepted voice recognition approaches used so far. The feature extraction in this study begins with transforming the speech signal into the frequency domain through the Fourier transform (FFT). A Discrete Fourier Transform (DFT) will then be used to process a framed and widowed signal to change each speech frame from the time domain into the frequency domain. The frequency power spectrum that results will be covered in accordance with the mel scale; hence, it will be changed into the mel spectrum. The cepstrum is then obtained by finding out the inverse Fourier transform of the logarithm of an approximated spectrum of a signal. Finally, after the log mel spectrum is obtained, the Discrete Cosine Transform (DCT) is used to change the spectrum back into the time domain. This method was used by the system to determine speakers with an accuracy of 80% in low noise environments and of 73.3% in moderate noise environments.

The work of [12] does not provide high accuracy in terms of matching the voice of users and identifying it under conditions of noise. Hence, two proposed speaker identification systems that used distinct classifiers, i.e.,

GMM and DNN, were examined. The accuracy of these two models was then determined to overcome this issue.

4. Noise Analysis Methodology

Two kinds of data sets were used to analyze the proposed systems, which were an open-source library known as Voxforge and prepared dataset. Voxforge is a collection of free speech and acoustic model repository used by open-source speech recognition engines. It includes 34 speakers and there are 10 voice clips for each speaker. There is no noise in these voice clips that are between 3 to 10 seconds long. The dataset was developed by asking a few volunteers to record their voice. The volunteers included 7 females and 4 males who registered 20 samples, each of which was 8 to 20 seconds long. The speakers ensured that they simulated reality by using the microphone of their headsets to record their voice. This is because the users of the proposed system will sign into systems like bank accounts through their mobile phones or laptops. This is considered as the foremost step in speech signal processing. A free and open-source digital audio editor and recording application software known as Audacity is used initially to eliminate noise in the dataset. This tool is available across different platforms, i.e., Windows, Linux, macOS and other Unix-like operating systems. MATLAB program is then used to add low noise and heavy noise to the registered audio. This provides us with three kinds of voice clips: clean clips without any noise, clips with low noise and finally, clips with heavy noise. The histograms of audio file following the addition low noise and heavy noise to clean voice are presented in Fig 1.

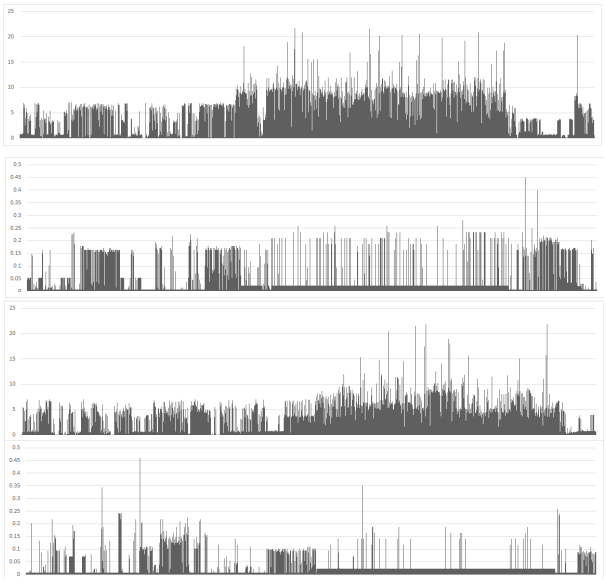

Fig. 1 Histogram of speaker voice with low and heavy noise

4.1 Speaker Identification based on MFCC

The pre-processing phase is followed by the assessment of the GMM-based voice recognition system as illustrated in Fig2. Initially, Voxforge dataset is employed for assessment of the system followed by the system assessment against the prepared dataset containing both clean and noisy voice clips. The 3 key modules involved in the functioning of GMM-based speaker identification system are: Feature extraction module, Modelization module and Decision module. Figure 4 offers an insight into the methodology of GMM and MFCC systems. In the Feature extraction module, the speech features are extracted from frames and shown in the form of vector. Usually, the feature extraction MFCC (Mel Frequency Cepstral Coefficients) technique is used. These coefficients are obtained on a twisted frequency scale on the basis of known human auditory perception. The method of estimating the MFCC coefficients is as follows: Initially, a speech signal is divided to get an overlapped frame. In the subsequent step, hamming or hamming window is used for application of windowing on the frames. Next, Fast Fourier Transform is employed to determine the frequency of signal. The subsequent step involves application of the Mel scale filter bank to the frame followed by the application of algorithm. In the last step, the discrete cosine transformation (DCT) is applied on the frame to determine coefficients [13].

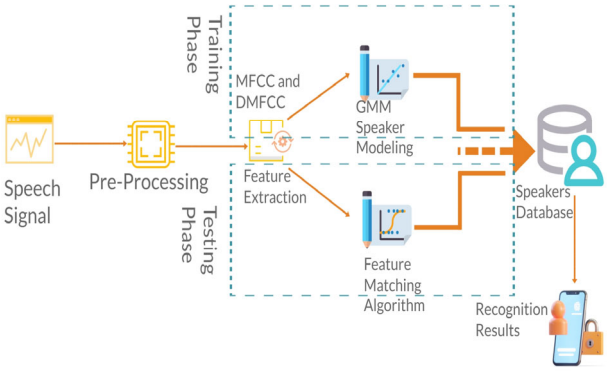

Fig. 2 Methodology of MFCC and GMM system

The summary of the steps involved in the process is given below:
The first step involves the transformation of windowed frame into the frequency domain through the use of Discrete Fourier Transform (DFT).3

$$X(k) = \sum_{n=0}^{N-1} x(n).w(n).exp^{(-j2\pi nk/N)} \quad (1)$$

In the above equation, k=0, 1…. N-1. Moreover, k indicates the frequency while w (n) indicates a time-window. In the

second step, the frequencies in the frame are subjected to 40 filters in the Mel filters bank.

$$X(m) = \sum_{k=0}^{N-1} |X(K)|.H(k,m) \quad (2)$$

In Eq. 2, m=1, 2…, M. In this equation, M indicates the number of filters included in Mel filter bank. It is important to understand that Mel filter bank H (m) includes triangular filters defined on the basis of Mel-frequencies. The Mel-frequencies are denoted by mel (f) and given as follows:

$$mel(f) = 2595.log\left(1 + \frac{f}{700}\right) \quad (3)$$

In the third step, the logarithm of the amplitude energies is computed.

$$\acute{X}(m) = \log\left(X(m)\right) \quad (4)$$

At the end of the third step, the inverse DFT is evaluated through Discrete Cosine Transform (DCT)

$$C(l) = \sum_{m=1}^{M} \acute{X}(m).cos\left[\frac{\pi l}{M}\ (m - \frac{1}{2})\right] \quad (5)$$

In the above equation, l=1, 2…, M and C (l) indicates the lth MFCC coefficient. This phase involves the extraction of 40 dimensional MFCC & delta MFCC (DMFCC) features. The speaker identification features are represented more clearly because of DMFCC features leading to higher accuracy. Consequently, the delta MFCC (DMFCC) is done on 20 MFCC features extracted from a voice clip. This is followed by development of 40 feature vector by combining DMFCC features. The training and pattern recognition processes in the subsequent module will make use of this feature vector as an input.

4.2 Modelization Features

The speech-based pattern recognition mainly applies Gaussian mixture model GMM which may be attributed to the ability of GMM to estimate various density distributions [14]. The weighted sum of all the component densities is referred to as the Gaussian mixture density. It is mathematically expressed as follows:

$$P(\bar{X})P(\bar{X}|\lambda) = \sum_{t=1}^{M} P_i\ B_i\ (\overline{X}) \quad (6)$$

In the above equation, X represents a random vector with N dimensions, $B_i(\bar{X}), i = 1 \ldots \ldots M$ represent various component densities, P_i indicates various mixture weights where i=1,2,3…….M . The component densities can be represented as N-variate Gaussian function as indicated below:

$$B_i(\vec{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}\ \sum|_{\vec{z}}^{2}} exp\left\{-\frac{1}{2}\left(\vec{x} - \overrightarrow{M_l}\right)\sum\left(\vec{x} - \overrightarrow{M_l}\right)\right\} \quad (7)$$

$\overrightarrow{M_l}$ indicates the mean vector while $\sum i$ indicates covariance matrix. The condition of $\sum_{i=1}^{M} P_i$ =1 is fulfilled by the mixture weights. The covariance matrices, mean vectors and mixture weights of all component densities collectively form the complete Gaussian mixture density as indicated below:

$$\lambda = \{P_i, M_i, \sum i\}, i = 1,2,3 \ldots \ldots M \quad (8)$$

A GMM model developed for each speaker λ is used to indicate each speaker in the Speaker identification system. The GMM models developed earlier are loaded in the testing phase. This is followed by recording input waveform of each registered speaker. Next, each speaker in the database is assigned a score computed on the basis of Maximum likelihood algorithm. After the assigning of scores, the GMM models depicting the highest score will be identified as the real speaker. Figure 4 gives an insight into the testing phase.

4.3 Speaker Identification based on DNN

In contrast to traditional GMM schemes, the deep neural network (DNN) is expected to yield greater accuracy. The higher accuracy depicted by DNN may be attributed to its multilevel distributed input representation which results in significantly more compact DNN in comparison to GMM. Moreover, use of DNN allows handling huge quantity of data without the need of any detailed assumptions about the input data or overtraining for obtaining a robust model [15]. Hence, DNN is a preferable option for Speaker Identification Systems.

In speaker identification systems, DNN acts as a classifier and helps to determine the speaker on the basis of his voice. Information in DNN travels in a single direction since it is fully-connected feed-forward neural network. The activation function of rectified linear units (ReLU) is usually used for hidden layers in DNN. Moreover, in the DNN, softmax layer serves the functions of an output layer. DNN derives back-propagation with the help of cross-entropy function.
Hence, considering level j, the input x_j is plotted as input of upper layer as corresponding activation y_i

$$y_j = ReLU\left(x_j\right) = \max\left(0, x_j\right) \quad (9)$$

$$x_j = b_j + \sum_i w_{ij}\ y_i \quad (10)$$

In the above equations, *i* represents hidden units in the lower layer while b_i represents bias associated with the unit *j*.

This is followed by the action of hidden units in mapping the class probability p_j from input x_j as indicated below:

$$p_j = \frac{\exp(x_j)}{\sum_l \exp(x_l)} \quad (11)$$

In equation (11), l indicates all the classes. Consequently, the training phase involves estimating the back-propagation gradients with application of cross-entropy function. This is indicated in the equation below:

$$C = -\sum_j t_j \, log(p_j) \quad (12)$$

In the above equation, t_j denotes the target probability of the class j. In case of true class, t_j=1 while in case of false class, t_j = 0.

Fig.3 gives an idea of the DNN network topology. The employed DNA network is evidently comprised of an input layer, an output layer and 6 hidden layers containing variable number of neurons.

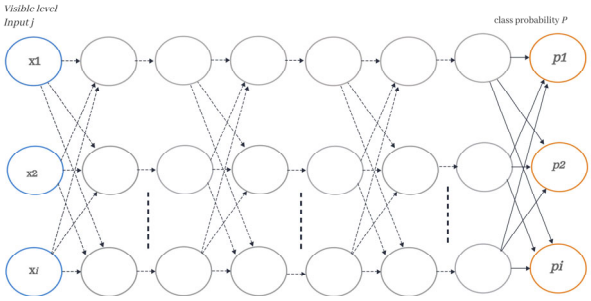

Fig3. The Six hidden layers DNN network topology

4.4 Feature Extraction Phase

The feature extraction process in this phase involves the use of the Librosa library which allows processing of audio files in python. Librosa library is helpful in resolving various audio classification problems. Hence, the NN-based speaker identification system will perform feature extraction through Librosa. The audio files of each speaker included in the Librosa library are classified in such a way that 20% data of each speaker is reserved for training the model. Hence, 14 files are used for training, 4 files for testing and 2 files for validation.

Initially, the training is performed by loading the audio files along with their corresponding label. This is followed by using data-frame for feature extraction. The process involves following the path of an audio file in computer to get access to it. Next, each data-frame row will experience execution of an iterative function. The feature extraction process involves the use of Mel-frequency Cepstral Coefficients (MFCCs), Mel-scaled Spectrogram, Chromagram, Spectral Contrast and Tonal Centroid Features (tonnetz). As result of feature extraction, we obtain a separate vector for each speaker based on 193 features as well as the corresponding label assigned to that vector.

5. Results and Discussion

While 11 speakers including 7 females and 4 males were registered in the developed dataset and included 20 voice clips in English language with 8 to 20 seconds duration, Percentage Identification Accuracy (PIA) of the voice recognition system can be determined using the equation given below:

$$PIA = \frac{\text{number of correct identifications}}{total\ number\ of\ trials} * 100\%$$

The Voxforge dataset included voice clips of 34 speakers. This dataset was used to test the performance of GMM-based voice recognition system. 5 out of the total voice clips were used in training phase while five others were used in testing phase. A total of 170 testing attempts were made.

Moreover, the developed dataset was used for assessment of the system both in the presence and absence of noise. The system was also assessed in the presence of both low and heavy noise. 15 voice clips from the developed dataset were used to train the system (in noise-free environment, low-noise environment and loud noise conditions). After this, the 5 voice clips of each of the 11 speakers kept for testing purposes are used to test the system accounting for a total of 55 testing attempts.

Table 1: PIA of Speaker Identification System based GMM and MFCC

Voxforge	***No- noise***	***Low-noise***	***Heavy-noise***
100%	100%	83.63%	38.18%

The testing revealed that GMM-based speaker identification system is affected by noise. In low-noise conditions, the system depicted a decline in the PIA from 100% to 83.63% and in case of heavy noise, the accuracy was found to be 38.18%.

Similarly, 14 voice-clips of each speaker registered in the developed dataset were used in training the NN-based speaker identification system while 4 clips of each speaker were used for assessing the system in the absence of noise

as well as in the presence of low noise and high noise. In the end, 2 voice clips of each speaker were used for validation purposes. A total of 44 testing attempts were made to reveal the outcomes depicted in table 2.

The training and validation accuracy with respect to epoch has been depicted in fig4. It is apparent from the figure that increasing epoch results in greater accuracy.

Table 2: PIA of Speaker Identification System based DNN

No- noise	*Low-noise*	*Low-noise*
97.7%	97.7%	93.2%

Table 2 clearly indicates that NN-based speaker identification system is not significantly affected by noise. The table also shows no change in accuracy in low noise conditions and slight decline in accuracy in the presence of heavy noise. Furthermore, data recorded in fig 5 shows that noise does not affect the NN-based speaker identification system; however, noise has some effect on the GMM-based speaker identification system which is evident from the decline in PIA.

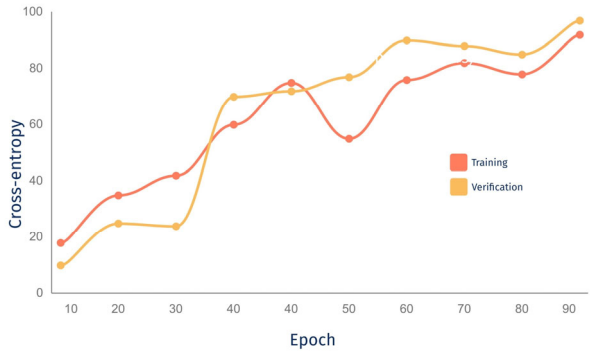

Fig4. Training and Validation Accuracy by Epoch in DNN model

A number of limitations have been found in voice-based identification systems; for instance, the use of voice-based identification system in banks is quite risky and can lead to huge fraudulent activities if an illegitimate individual manages to get access to real user's account.

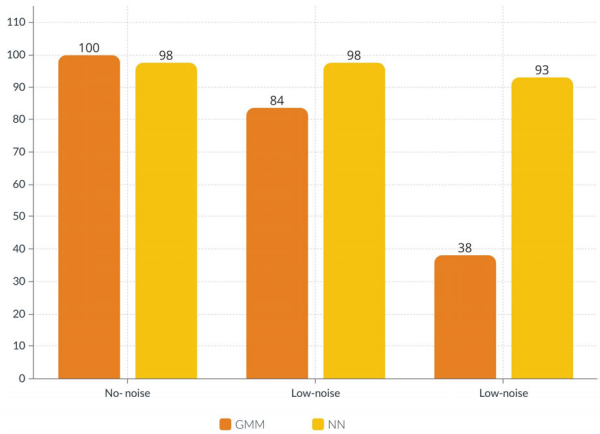

Fig5. Accuracy of Models under different noise circumstances on prepared dataset

Moreover, since these systems use ML, the system can only identify from among the fed options with no ability to identify a new user. In such conditions, if the system receives the voice of a new user, it will classify the new user by assessing its features and comparing them with the features of already registered users and grant him access in case of similarity in features which is likely to provide the new user with an access to someone else's account. The accuracy of the identification system can decline in the presence of heavy noise like rain, echo or loud environmental noise resulting in inaccurate user identification. In addition, bank accounts can be easily accessed by someone having a voice-clip of the real account-holder. It is recommended to employ the authentication aspects like password, fingerprint and face print for further authentication of the individual. Consequently, the system is expected to depict higher PIA and more accurate identification. Also, it is recommended for accurate identification to ensure that the training set is free from noise

6. Conclusion

The main function of the Speaker identification system is to identify individuals on the basis of their voices. These systems help to improve the efficiency of the identification systems by making them more secure. Security of identification systems is usually enhanced by using the biometric like fingerprint and voiceprint since these are unique for each individual and allow easy identification. But, the efficiency of speech identification systems can be adversely affected by environmental noise. Therefore, it is imperative to check the impact of noise on voice-based identification systems. For this purpose, we will be using two different classifiers of GMM and DNN for testing the voice-based identification systems. In the testing process,

the MFCC-based as well as GMM-based Speaker identification systems were operated under different noise conditions to check their performance in the presence of environmental noise. Two datasets were used in the process; the first was Voxforge while the second had been developed for testing purposes. The testing results indicated a decline in the system accuracy (from 100% to 38.18%) under the impact of heavy noise. Moreover, the developed dataset was used to analyze the performance of DNN-based Speaker identification system. The system was subjected to low noise as well as high noise to evaluate the impact of noise. The testing depicted efficient functioning of the system even in the presence of noise. The system depicted 93.2% accuracy in the presence of heavy noise. Hence, it is concluded that DNN-based identification systems perform better even in the presence of noise. The future research must focus on enhancing the efficiency of the speaker identification systems and reducing the impact of noise by using larger datasets for training the systems.

References

[1] Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). Introduction to biometrics. Springer Science & Business Media.

[2] Ahmed, A. A. E., & Traore, I. (2005, June). Anomaly intrusion detection based on biometrics. In Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop (pp. 452-453). IEEE Huerta-Canepa, G., Lee, D.: A virtual cloud computing provider for mobile devices. In: Proc. of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond (2010)

[3] Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv preprint arXiv:1003.4083.

[4] Melin, P., Urias, J., Solano, D., Soto, M., Lopez, M., & Castillo, O. (2006). Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithms. Engineering Letters, 13(3).

[5] Kumar, G. S., Raju, K. P., CPVNJ, M. R., & Satheesh, P. (2010). Speaker recognition using GMM. International Journal of Engineering Science and Technology, 2(6), 2428-2436.

[6] Snyder, D., Ghahremani, P., Povey, D., Garcia-Romero, D., Carmiel, Y., & Khudanpur, S. (2016, December). Deep neural network-based speaker embeddings for end-to-end speaker verification. In 2016 IEEE Spoken Language Technology Workshop (SLT) (pp. 165-170). IEEE.

[7] Bird, S., Boguraev, B., Kay, M., McDonald, D., Hindle, D., & Wilks, Y. (1997). Survey of the state of the art in human language technology (Vol. 13). Cambridge university press.

[8] Yanushkevich, S., Howells, G., Crockett, K., O'Shea, J., Oliveira, H. C. R., Guest, R., & Shmerko, V. (2019, October). Cognitive Identity Management: Risks, Trust and Decisions using Heterogeneous Sources. In Proceedings of the First IEEE International Conference on Cognitive Machine Intelligence.

[9] Kounoudes, A., Kekatos, V., & Mavromoustakos, S. (2006, April). Voice biometric authentication for enhancing Internet service security. In 2006 2nd International Conference on Information & Communication Technologies (Vol. 1, pp. 1020-1025). IEEE.

[10] Boles, A., & Rad, P. (2017, June). Voice biometrics: Deep learning-based voiceprint authentication system. In 2017 12th System of Systems Engineering Conference (SoSE) (pp. 1-6). IEEE.

[11] Hintze, D., Füller, M., Scholz, S., Findling, R. D., Muaaz, M., Kapfer, P., ... & Mayrhofer, R. (2019). CORMORANT: On Implementing Risk-Aware Multi-Modal Biometric Cross-Device Authentication For Android.

[12] Kumar, A. N. A., & Muthukumaraswamy, S. A. (2017, April). Text dependent voice recognition system using MFCC and VQ for security applications. In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 2, pp. 130-136). IEEE.

[13] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232.

[14] Beal, M. J., Ghahramani, Z., & Rasmussen, C. E. (2002). The infinite hidden Markov model. In Advances in neural information processing systems (pp. 577-584).

[15] Yu, D., Deng, L., & Dahl, G. (2010, December). Roles of pre-training and fine-tuning in context-dependent DBN-HMMs for real-world speech recognition. In Proc. NIPS Workshop on Deep Learning and Unsupervised Feature Learning.

Wahid Rajeh received the B.S. in Computer Science in 2007 from Taiba University, Medina, and M.S. degrees in Information Technology in 2010 form Queensland University of Technology Brisbane. He received the PhD degree in Computer Applied Technology in 2018. He is now an assistant professor since at College of computer and Information Technology at University of Tabuk. His research interest includes Cloud Computing – Big Data – Security Analysis – Grid Computing – Risk Assessment.