❒   782

Comparison of feature extraction and normalization methods for speaker recognition using grid-audiovisual database

Musab T. S. Al-Kaltakchi[1], Haithem Abd Al-Raheem Taha[2], Mohanad Abd Shehab[3], Mohammed A. M. Abdullah[4]

[1,2,3]Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq
[4]Department of Computer and Information Engineering, Ninevah University, Mosul, Iraq

Article Info

ABSTRACT

In this paper, different feature extraction and feature normalization methods are investigated for speaker recognition. With a view to give a good representation of acoustic speech signals, Power Normalized Cepstral Coefficients (PNCCs) and Mel Frequency Cepstral Coefficients (MFCCs) are employed for feature extraction. Then, to mitigate the effect of linear channel, Cepstral Mean-Variance Normalization (CMVN) and feature warping are utilized. The current paper investigates Text-independent speaker identification system by using 16 coefficients from both the MFCCs and PNCCs features. Eight different speakers are selected from the GRID-Audiovisual database with two females and six males. The speakers are modeled using the coupling between the Universal Background Model and Gaussian Mixture Models (GMM-UBM) in order to get a fast scoring technique and better performance. The system shows 100% in terms of speaker identification accuracy. The results illustrated that PNCCs features have better performance compared to the MFCCs features to identify females compared to male speakers. Furthermore, feature wrapping reported better performance compared to the CMVN method.

Corresponding Author:

Musab T. S. Al-Kaltakch
Department of Electrical Engineering,
Mustansiriyah University, Baghdad, Iraq.
Email: musab.tahseen@gmail.com

1. INTRODUCTION

Biometrics characteristics can be divided into two main types [1]: Behavioural and Physiological traits. Examples of the behavioral traits include Voice, Signature, Gait, and Keystroke [2-5], and [6]. On the other hand, Physiological traits include Iris, Retina, Face, Ear, DNA, Hand Geometry, Palm, and Fingerprint [7-12]. Speaker recognition is a form of behavioral biometrics which is used to verify an individual's claimed identity from his or her voice. Generally speaking, a speaker recognition system works in two modes: verification or identification. In the verification mode, the system is deciding if a speaker is a particular person or is among a group of individuals [13]. On the other hand, the system which identifies who is speaking is called a speaker identification system. Figure 1 illustrates the block diagram of a speaker recognition system where two operational stages are used, namely: training and testing. In the training stage, the speech signals from all speakers are obtained in order to build the speaker model. Basically, the training phase is constructed off-line while in testing the actual operation of the system is achieved (on-line) where the speech from an unknown speaker is compared with each of the trained speaker models [13] to identify a speaker.

Speaker recognition has enormous applications some of these are [13-14]: Control access to services such as mobile banking; remote access to computers; voice mail; security control of confidential information area. In order to build a robust speaker recognition system, the effect of feature extraction method should be

investigated. Also, as the acoustic signal has a different characteristic which varies from male to female, finding a suitable feature extraction method that works with all these variations is important.

In this paper, we investigate a method for text-independent speaker recognition (closed-set speaker identification) by using Gaussian mixture models with the Universal Background Model (GMM-UBM). This work is conducted by employing the GRID-Audiovisual database [15]. Then, based on a person's voice, identification can be established. In addition, diverse feature extraction and feature normalization methods are employed such as the Mel Frequency Cepstral Coefficients (MFCCs) and the Power Normalized Cepstral Coefficients (PNCCs). Also, feature normalization is applied using the CMVN and feature warping to attain a good comparison between features for this task. In addition, two comparisons are made based on two experiments based on the feature extraction and normalization methods as shown in Experiments (1) and (2) later in this paper.

This paper is organized as follow: Section 1 introduces the speaker recognition and includes the main applications. Section 2 Introduces the proposed method along with the employed feature extraction and normalization methods. The experiments and results will be covered in Section 3. Finally, Section 4 concludes this paper.
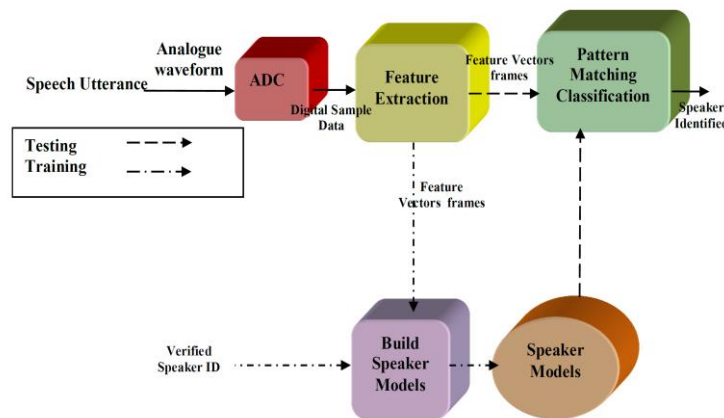

Figure 1. The Block Diagram for a Speaker Recognition System [13]

2. PROPOSED METHOD

Figure 2, shows the main block diagram for the GMM-UBM system which is used in this work for a speaker identification system. The details of the block diagram are explained in the next subsections.

2.1. Feature Extraction and Feature Normalization

Feature extraction converts the original speech recordings into a data set with a reduced number of variables which contain the most significant information [13, 16]. Feature extraction is performed to remove the unwanted information or to reduce the cost and thereby reduce the system complexity in order to acquire better performance [13, 17]. Figure 3 illustrates the main differences between the two features extraction methods used in this paper: MFCC and PNCC [18, 19]. In addition, this work presents two feature normalization methods in order to alleviate the channel effects by either Gussianize the feature distribution (applying the Feature Warping) or by taking the mean and variance of the Cepstrum features (employing the CMVN) [20].

2.2. Mel Frequency Cepstral Coefficients (MFCCs)

In 1963, Bogert et al. fabricated the term "Cepstrum", and this term comes from reversing the first syllable of word spectrum. The Cepstrum is the inverse Fourier transform of the log spectrum. According to Figure 3, the implementation of MFCCs features can be classified into five sections [18, 21] namely: 1) Pre-emphasis 2) Frame blocking and windowing 3) Fast Fourier Transform (FFT) 4) Mel-scaled filter bank and 5) Generate MFCCs features.

In the first section, the pre-emphasis is a first-order Finite Impulse Response (FIR) filter which is utilized to compensate the high-frequency components which were suppressed during the human voice production. In the second section, framing is used to treat the non-stationary behavior for the speech signal while Hamming window is employed to mitigate the discontinuities at the edges of the speech signals. Then, FFT is applied at the third section and thereby the spectrum is converted to Mel-scale in the fourth section. In

the Mel-scale, the behavior is linear frequency spacing below 1,000 Hz and with logarithmic spacing over 1 kHz. In the final section, the log for Mel-spectrum is used and transferred back to the time domain to produce the MFCCs features [22]. The reader can refer to [18, 21] for further information. The bandwidth and spacing are calculated by a constant interval of Mel-frequency [21] as shown in (1):
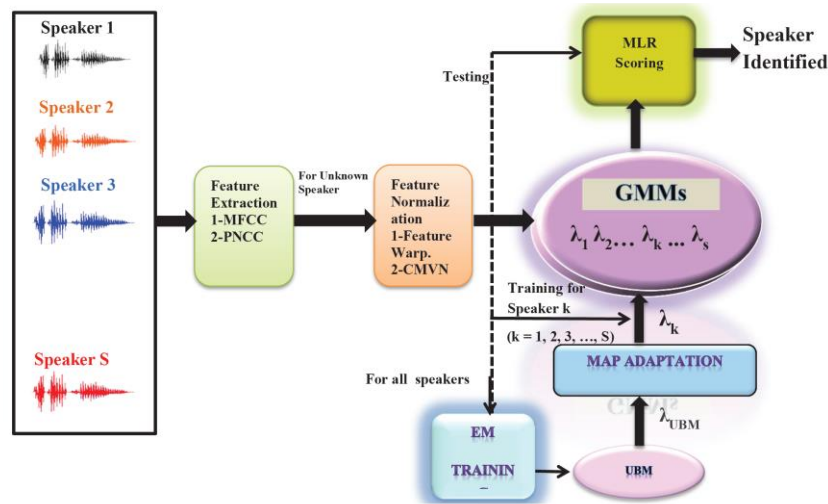

Figure 2. The GMM-UBM block diagram [13]

$$Mel(f) = \begin{cases} f & if\ f \leqslant 1kHz \\ 2595\ log_{10}\left(1 + \frac{f}{700}\right) & if\ f > 1kHz \end{cases}$$

(1)

The MFCC is determined via the following equation [20]:

$$C_n = \sum_{i=1}^{K}\left(log_{10}\ Y(i)\ cos\left[i\ \frac{2\pi n}{N}\right]\right)$$

(2)

where: Cn is the Cepstrum coefficients, n = 1, 2,..., L, L is the cepstral coefficients, N is the number of FFT points (N = 512), K is the number of channel filter banks (K = 40), and Y(i) is the output of the ith filter bank.

2.3. Power Normalized Cepstral Coefficients (PNCCs)

There are three main stages that are used to compute the PNCCs features and they are [19]: 1) Initial processing 2) Environmental processing and 3) Final processing. In the initial stage, the pre-emphasis filter is applied followed by the Short Time Fourier Transform (STFT). In addition, the Gammatone Filter Bank (GFB) is employed instead of triangular filter bank which used in the MFCCs features. However, the environmental stage consists of temporal processing and spectral smoothing which both affect the system accuracy under white noise. Furthermore, medium time power is used in order to estimate then compensate for the noise. Moreover, to mitigate the noise effect, asymmetric noise suppression is used in order to subtract the spectrum of the noise level. Finally, the Discrete Cosine Transform (DCT), as well as the mean normalization, are used to determine the PNCC features [23]. Further details for PNCCs features are provided by [19]. The DCT and the mean normalization is shown in (3).

$$\mu[m] = \lambda_\mu\ \mu(m-1) + \frac{(1-\lambda_\mu)}{L}\sum_{l=0}^{L-1}T[m,l]$$

(3)

where: l is the channel incident, m represents the frame incident, L is the number of frequency channels, and λ_c is the forgetting factor which is equal to 0.999, whereas the T[m,l] represents the time-frequency normalization.
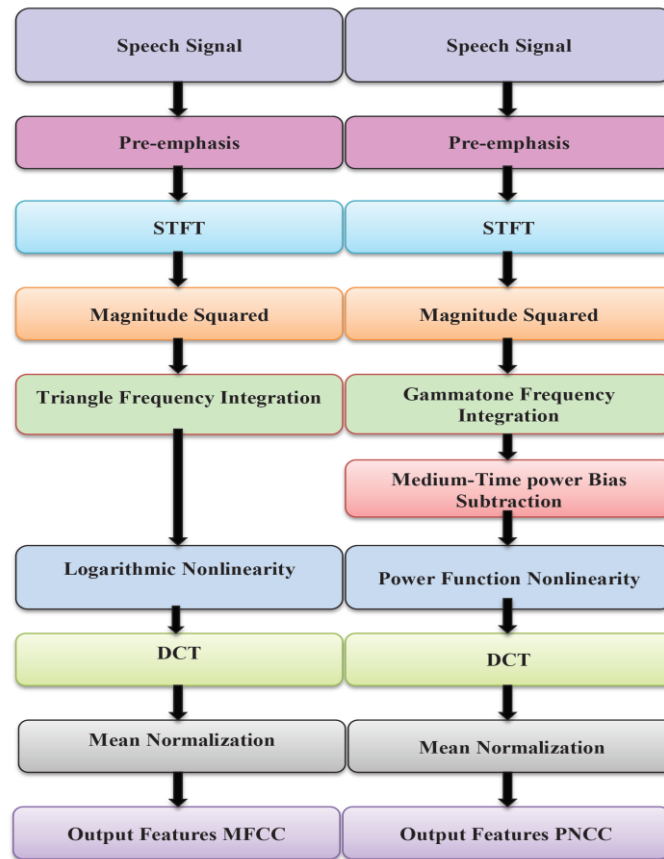
Figure 3. Differences Between the MFCCs and the PNCCs Features [19]

The power-law nonlinearity is produced by a value of 1/15 which is empirically chosen to give acceptable accuracy in white noise and without any significant impact on recognition accuracy in clean speech, as shown in (4) [19]:

$$V[m, l] = U[m, l]^{\frac{1}{15}}$$

(4)

where U[m,l]: is the normalized power. Table 1 summarizes the difference between the MFCCs and PNCCs features. On the other hand, the difference between feature warping and CMVN are shown in Table 2.

3.   EXPERIMENTAL RESULTS AND DISCUSSION

In this section, two experiments are conducted on the speech sample from the GRID-Audiovisual database where all files are converted from video to audio prior to use. In addition, each speaker is represented by one model, then in this experiment, eight speech files are used for training to represent eight speakers (8 speaker models). Furthermore, each speaker model has a length between 2-3 minutes. In addition, in the testing stage, three speech files are selected from each speaker (3 tests files/ speaker) to result in 24 tests. Similarly, each test file has the same length as the training file.

Table 1. Comparisons Between the MFCC and the PNCC Features [14, 24]

MFCC	PNCC
Mel Filter Bank (Triangular)	Gammatone Filter Bank
Logarithmic Non-Linearity	Power Law Non-Linearity
Less Accurate	Better Accuracy in the Presence of White Noise
Less complex	33% more complex than MFCC

Table 2. Comparisons Between the Feature Warping and the CMVN [24-25]

Feature Warping	CMVN
Aim: Mitigate the linear channel effect.	Aim: Remove the linear channel effect.
The middle frame of window is normalized based on the rank in the array of the sorted feature values.	The middle frame of window is normalized based on the computed mean and variance.
The overall feature distribution is wrapped to the standard normal distribution.	Feature stream is mapped to the standard normal distribution.

Table 3. The Scoring of GMM-UBM for the Feature Warping-MFCCs Approach for 8 Models and 24 Tests

Scoring for GMM-UBM Approach [8Model,24Test] Using Feature Warping of MFCCs											
Model	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
1	0.888	0.891	0.659	-0.052	0.235	-0.299	-0.098	-0.281	-0.002	-0.569	-0.415	-1.026
2	-0.544	-0.458	-0.328	1.105	1.032	1.025	-0.678	-0.530	-0.511	-1.057	-0.755	-1.368
3	-0.703	-0.312	-0.202	-0.583	-0.842	-0.966	1.125	0.964	0.889	-1.124	-0.592	-1.022
4	-1.357	-1.396	-1.677	-0.944	-1.505	-1.361	-1.138	-1.188	-1.167	0.658	0.741	0.756
5	-1.013	-1.105	-0.704	-0.744	-0.853	-0.762	-0.514	-0.236	-0.684	-0.816	-0.896	-0.960
6	-0.692	-0.644	-0.353	-1.224	-1.209	-0.897	-0.997	-0.407	-0.546	-0.814	-0.781	-0.758
7	-1.445	-1.546	-1.916	-1.648	-2.098	-1.910	-2.225	-2.502	-2.024	-1.334	-1.359	-0.978
8	-1.682	-1.634	-1.743	-2.317	-2.072	-1.876	-2.316	-2.404	-2.273	-0.924	-1.199	-0.758
Speaker ID	Male 1	Male 1	Male 1	Male 2	Male 2	Male 2	Male 3	Male 3	Male 3	Male 4	Male 4	Male 4
Model	T13	T14	T15	T16	T17	T18	T19	T20	T21	T22	T23	T24
1	-0.715	-1.001	-0.928	-0.429	-0.299	-0.520	-1.184	-1.221	-1.073	-2.029	-1.681	-1.821
2	-0.758	-1.036	-1.087	-1.1008	-0.718	-0.957	-1.618	-1.649	-1.624	-1.969	-2.102	-2.102
3	-0.844	-1.164	-0.591	-0.486	-1.071	-0.908	-1.699	-2.662	-2.221	-2.319	-2.312	-2.118
4	-1.047	-0.319	-1.250	-0.550	-1.104	-0.632	-1.349	-1.281	-1.322	-0.332	-0.902	-0.693
5	0.754	0.441	0.664	-0.438	-0.605	-0.532	-0.888	-1.862	-1.639	-0.944	-1.014	-0.559
6	-0.359	-0.630	-0.341	0.279	0.287	0.583	-0.923	-1.337	-1.215	-1.290	-1.192	-0.808
7	-1.568	-1.287	-1.421	-1.376	-1.377	-1.551	0.777	1.750	1.044	-0.389	0.002	-0.662
8	-1.453	-0.836	-0.924	-1.345	-1.152	-1.080	-0.280	-0.172	-0.328	1.183	1.0229	1.131
Speaker ID	Male 5	Male 5	Male 5	Male 6	Male 6	Male 6	Female7	Female 7	Female 7	Female 8	Female 8	Female 8

Table 4. The Scoring of GMM-UBM for the Feature Warping-PNCCs Approach for 8 Models and 24 Tests

Scoring for GMM-UBM Approach [8Model,24Test] Using Feature Warping of PNCCs											
Model	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
1	1.033	0.681	0.857	-0.053	-0.074	-0.435	-0.338	-0.465	-0.228	-0.441	-0.830	-0.439
2	-0.865	-0.683	-0.439	1.082	1.169	1.122	-0.671	-0.992	-0.763	-1.259	-0.399	-1.164
3	-0.755	-0.534	-0.540	-1.171	-1.413	-1.430	0.9007	0.904	0.867	-0.759	-1.159	-1.298
4	-2.470	-1.692	-2.619	-0.596	-2.043	-1.710	-1.165	-1.179	-0.405	0.512	0.850	0.438
5	-1.189	-1.673	-0.848	-0.930	-1.044	-0.609	-0.385	-0.131	-0.611	-1.074	-0.963	-0.518
6	-0.864	-0.993	-0.449	-2.170	-0.954	-1.052	-1.277	-0.540	-0.747	-1.091	-1.205	-1.376
7	-2.11	-2.653	-2.997	-2.757	-3.241	-2.646	-3.236	-3.863	-3.332	-2.719	-2.735	-2.185
8	-1.743	-1.552	-2.045	-2.255	-1.675	-1.463	-2.143	-2.063	-1.945	-1.230	-1.594	-1.044
Speaker ID	Male 1	Male 1	Male 1	Male 2	Male 2	Male 2	Male 3	Male 3	Male 3	Male 4	Male 4	Male 4
Model	T13	T14	T15	T16	T17	T18	T19	T20	T21	T22	T23	T24
1	-0.688	-0.740	-0.573	-0.552	-0.192	-0.449	-1.235	-1.132	-1.291	-1.542	-1.356	-1.175
2	-0.785	-0.748	-0.729	-0.867	-0.428	-0.911	-1.809	-2.076	-2.485	-2.360	-2.307	-1.698
3	-1.643	-1.156	-0.606	-1.076	-1.752	-0.930	-2.754	-3.124	-3.155	-2.926	-2.412	-2.142
4	-1.542	-0.252	-1.170	-1.147	-1.693	-0.585	-1.808	-1.876	-1.699	-1.440	-1.172	-1.108
5	0.672	0.525	0.558	-0.577	-0.677	-0.669	-1.318	-1.975	-1.874	-1.845	-1.272	-0.904
6	-0.632	-1.279	-0.469	0.331	0.088	0.680	-1.573	-1.734	-1.519	-1.708	-1.737	-1.094
7	-2.846	-2.300	-2.545	-2.392	-2.415	-2.669	1.5309	2.063	1.878	-1.145	-1.123	-2.445
8	-1.128	-1.110	-1.179	-1.514	-1.581	-1.314	-0.545	-0.451	-0.900	1.286	1.202	1.005
Speaker ID	Male 5	Male 5	Male 5	Male 6	Male 6	Male 6	Female 7	Female 7	Female 7	Female 8	Female 8	Female 8

In Experiment (1): 13 coefficients are extracted by using the MFCCs and PNCCs features including zero coefficients for 8 different speakers (6 male, 2 female) from the GRID-Audiovisual database. Then, feature normalizations methods are employed (CMVN and feature warping) to the feature vectors of MFCCs and PNCCs. Figure 4, shows a comparison between the CMVN and the feature warping for both the MFCCs and the PNCCs features by using Mesh plot. Furthermore, Figure 4 Part (a) and Part (b) compare the CMVN and the feature warping for MFCCs features. Similarly, Part (c) and Part (d) compare the PNCCs features using feature wrapping and CMVN normalization. It is obvious that the coefficients have more significant values when using the feature warping rather than using the CMVN method for both the MFCCs and the PNCCs features. In addition, from the comparison of parts (a),(b),(c) and (d) in Figure 4, it is clear that the PNCCs features have higher amplitude compared with the MFCCs features for both the CMVN and the feature warping. It can also be seen that the feature wrapping method achieves better results compared with

the CMVN. Therefore, only feature warping is taken as a feature normalization for both feature extraction methods in the Experiment (2).
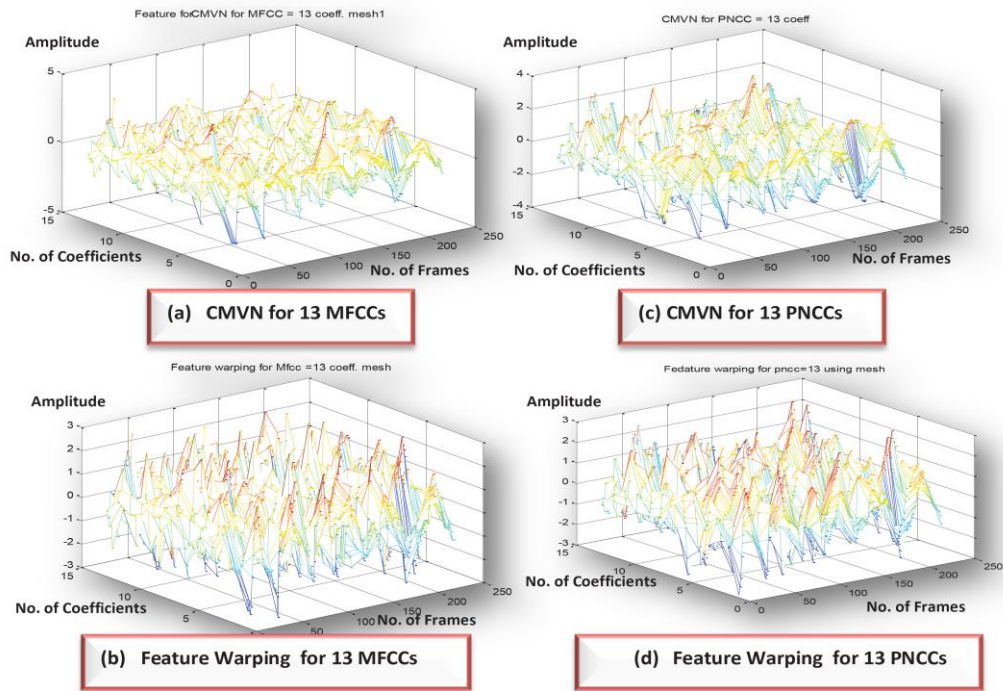

Figure 4. Comparison Between CMVN and Feature Warping for MFCCs and PNCCs Using Mesh Plot (a): CMVN for 13MFCCs (b): Feature warping for 13 MFCCs. (c): CMVN for 13PNCCs (d): Feature warping for 13PNCCs

In Experiment (2): the feature warping for MFCCs and PNCCs are used and the features are modeled by GMM-UBM approach. So, the trials between modeling and testing are 192 trials (8Model, 24Test) which yield 192 scores. The setting parameters used in this experiment are 16 MFCCs and PNCCs coefficients, 16 GMCs and 20 of expectation maximization iterations, and the MAP adaptation relevance factor is 10. Table 3 shows, the scoring for GMM-UBM for 8 models and 24 tests for the Feature Warping - MFCCs approach. On the other hand, Table 4 shows, the scoring for GMM-UBM for 8 models and 24 tests for the Feature Warping-PNCCs approach. The results from each table can illustrate the scoring between eight different speaker models against 24 Tests from all speakers (3Test/speaker). According to this experiment, the tests (T1, T2, T3) belong to the speaker 1 (Model1) and the tests (T4, T5, T6) belong to the speaker2 (Model 2) and tests (T22 T23, T24) belong to the speaker8. According to Table 3 and Table 4, the scores for speaker 1 as an example can be explained by the first row which represents the scoring between speaker model 1 against all the 24 tests. It is clear that the maximum scores can be taken from the first three tests, likewise for other speakers.

It is obvious that all the highlighted scores represent the maximum scores for each speaker (positive values), while all other scores in the same row have negative scores. In addition, all the tests are succeeded for the identification process and this yield 100% as identification accuracy or what is called the identification rate. As a comparison between the two tables, it seems that the system that uses the PNCCs features is more robust to identify females. On the other hand, empirical results show that the MFCC feature is more robust for male identification.

4.   CONCLUSION

In this work, two comparisons based on feature extraction and feature normalization methods were conducted on 8 speakers from the Grid-Audiovisual database through two main experiments for speaker identification task. Both systems were succeeded to identify the speech sample which resulted in an

identification rate of 100%. According to the experiments it was found that the system which employed the PNCCs features was more powerful to identify female and have the highest score (Maximum Log-Likelihood Ratio (MLLR) for females) compared with the corresponding system that used MFCCs features as shown in tests (T19, T20, T21, T22, T23, T24). On the contrast, the system that employed the MFCCs features seemed to identify male speakers better than female and reported the highest scores in most tests (MLLR) as shown in tests (T1 -T18). In terms of feature normalization, feature warping achieved better results for both MFCCs and PNCCs feature vectors compared with CMVN method.

REFERENCES

[1] K. Jain, A. A. Ross, and K. Nandakumar, "Introduction to Biometrics". *Springer Publishing Company, Incorporated*, 2011.

[2] R. R. Al-Nima *et al.,* "Personal verification based on multi-spectral finger texture lighting images," *IET Signal Processing*, vol. 12, no. 10, pp. 1154-1164, 2018.

[3] S. Singh, "The role of speech technology in biometrics, forensics and man-machine interface," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 9, no. 1, pp. 281, 2019.

[4] J. O. Pinz´on-Arenas, R. Jimenez-Moreno, and C. G. Pachón-Suescún, "Offline signature verification using DAG-CNN." *International Journal of Electrical & Computer Engineering (IJECE)*, vol. 9, pp. 2088- 8708, 2019.

[5] Acien, J. Hernandez-Ortega, A. Morales, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia, "*On the analysis of keystroke recognition performance based on proprietary passwords*," in 8th International Conference of Pattern Recognition Systems (ICPRS 2017), July 2017, pp. 1-6.

[6] M. Balazia and K. N. Plataniotis, "Human gait recognition from motion capture data in signature poses," *IET Biometrics*, vol. 6, no. 2, pp. 129-137, 2017.

[7] M. A. M. Abdullah, R. R. Al-Nima, S. S. Dlay, W. L. Woo, and J. A. Chambers, "Cross-Spectral Iris Matching for Surveillance Applications," In: Surveillance in Action, Advanced Sciences and Technologies for Security Applications. Cham, Switzerland: *Springer*; 2018. pp. 105-125.

[8] N. Hezil and A. Boukrouche, "Multimodal biometric recognition using human ear and palmprint," *IET Biometrics*, vol. 6, no. 5, pp. 351-359, 2017.

[9] R. P. Persada, S. Aulia et al., "Automatic face and VLP's recognition for smart parking system." *Telkomnika (Telecommunication, Computing, Electronics and Control),* vol. 17, no. 4, 2019.

[10] H. Sinha and P. K. Ajmera, "Upgrading security and protection in ear biometrics," *IET Biometrics*, vol. 8, no. 4, pp. 259-266, 2019.

[11] Chatterjee, P. Singh, V. Bhatia, and S. Prakash, "Ear biometrics recognition using laser biospeckled fringe projection profilometry," *Optics and Laser Technology*, vol. 112, pp. 368-378, 2019.

[12] W. Yang, S. Wang, J. Hu, G. Zheng, and C. Valli, "Security and accuracy of fingerprint-based biometrics: A review," *Symmetry*, vol. 11, no. 2, 2019.

[13] R. Togneri and D. Pullella, "An overview of speaker identification: Accuracy and robustness issues," *IEEE circuits and systems magazine*, vol. 11, no. 2, pp. 23-61, 2011.

[14] M. T. Al-Kaltakchi, W. L. Woo, S. Dlay, and J. A. Chambers, "Evaluation of a speaker identification system with and without fusion using three databases in the presence of noise and handset effects," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, p. 80, 2017.

[15] M. Cooke, J. Barker, S. Cunningham and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition". *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2421-2424, 2006.

[16] Rashed and W. M. Bahgat, "Modified technique for speaker recognition using ANN," *International Journal of Computer Science and Network Security*, vol. 13, no. 8, pp. 8-13, 2013.

[17] R. R. O. Al-Nima, M. A. Abdullah, M. T. Al-Kaltakchi, S. S. Dlay, W. L. Woo, and J. A. Chambers, "Finger texture biometric verification exploiting multi-scale sobel angles local binary pattern features and score-based fusion," *Digital Signal Processing*, vol. 70, pp. 178-189, 2017.

[18] Maesa, F. Garzia, M. Scarpiniti, and R. Cusani, "Text independent automatic speaker recognition system using MEL-frequency Cepstrum coefficient and gaussian mixture models," *Journal of Information Security*, vol. 3, no. 8, p. 335, 2012.

[19] E. Ambikairajah, J. M. K. Kua, V. Sethu, and H. Li, "*PNCC-ivector-SRC based speaker verification*," in Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference. IEEE, 2012, pp. 1-7.

[20] H. Beigi, "Signal Enhancement and Compensation". Boston, MA: *Springer US*, 2011, pp. 561-586.

[21] G. Nijhawan and M. Soni, "A new design approach for speaker recognition using MFCC and VAD," *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, vol. 9, no. 5, pp. 43-49, 2013.

[22] L. Shi, I. Ahmad, Y. He, and K. Chang, "Hidden Markov model based drone sound recognition using MFCC technique in practical noisy environments," *Journal of Communications and Networks*, vol. 20, no. 5, pp. 509-518, 2018.

[23] Kim and R. M. Stern, "Power-normalized cepstral coefficients (PNCCs) for robust speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 7, pp. 1315-1329, July 2016.

[24] M. T. S. Al-Kaltakchi, W. L. Woo, S. S. Dlay, and J. A. Chambers, "*Study of fusion strategies and exploiting the combination of MFCC and PNCC features for robust biometric speaker identification*," in 2016 4th International Conference on Biometrics and Forensics (IWBF), March 2016, pp. 1-6.
[25] M. T. S. Al-Kaltakchi, "Robust text independent closed set speaker identification systems and their evaluation," PhD Thesis, Newcastle University, 2018.

BIOGRAPHIES OF AUTHORS

Musab T. S. Al-Kaltakchi obtained the BSc in Electrical Engineering and MSc in Communication and Electronics from Mustansiriyah University, Iraq. He was awarded a PhD degree in Electrical Engineering-Digital Signal Processing from Newcastle University, UK in 2018. Currently he is a lecturer at the Electrical Engineering Department, Mustansiriyah University, Baghdad, Iraq. His research interests include Speaker identification and verification, Speech and audio signal processing, Machine learning, Pattern recognition, and Biometrics. Dr. Al-Kaltakchi is a member in the Institute of Electrical and Electronic Engineering (IEEE) and the Institute of Engineering and Technology (IET).

Haithem Abd Al-Raheem Taha received the B.Sc. degree in Electrical and Electronic Engineering, from Sarajevo, Yugoslavia in 1983 and M.Sc. degree in Electrical and Electronic Engineering, from Belgrade in 1985. He is currently a lecturer in the Electrical Engineering department, Mustansiriyah University, Iraq. His research interests include air crafting, image and signal processing.

Mohanad Abd Shehab received the B.Sc. and M.Sc. degrees in Electrical and Electronic Engineering, from the University of Technology, Iraq, in 1996 and 1998 respectively, and Ph.D. degree in Electronic Engineering, from Yildiz Technical University, Turkey, in 2018. He is currently a lecturer in the Electrical Engineering department, Mustansyriah University, Iraq. His research interests include image and signal processing, ANN, pattern recognition and machine learning.
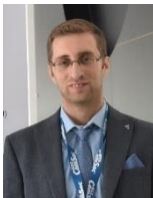
Mohammed A. M. Abdullah received the PhD degree in 2017 from the school of Electrical and Electronic Engineering at Newcastle University, UK. He received the BSc and MSc degrees in Computer Engineering in 2008 and 2010, respectively from the university of Mosul, Iraq. During 2018, he worked as a post-doctorate researcher in the department of engineering at the University of Leicester, UK. He is currently a lecturer with the Department of Computer and Information Engineering at Ninevah University, Iraq. His research interests are in the fields of pattern recognition, machine learning, signal processing and wireless communication. Dr. Abdullah is a member of the Institute of Electrical and Electronic Engineering (IEEE) and International Biometric Society (IBS). He is also an associate fellow of the UK Higher Education Academy (AFHEA).