Multimed Tools Appl (2019) 78:1569–1582
https://doi.org/10.1007/s11042-018-6258-0

Person authentication using speech as a biometric against play back attacks

A. Revathi[1] **· C. Jeyalakshmi**[2] **· K. Thenmozhi**[1]

Received: 23 December 2017 / Revised: 3 June 2018 / Accepted: 6 June 2018 /
Published online: 26 June 2018

Abstract This work presents the modules for authenticating the persons by using speech as a biometric against recorded playback attacks. It involves the implementation of feature extraction, modeling technique and testing procedure for authenticating the persons. Playback attacks are simulated by recording the original speech utterances using the speakers and mikes in a laptop using Audacity software. This work mainly involves the process for distinguishing original and recorded speeches and authenticating the speakers based on voice as a biometric. Features extracted from the original and recorded speeches are used to develop models for them. Voice passwords are assigned to the speakers and features are extracted from the training speech created by fusing the password specific original speech utterances. These features are applied to the training algorithm to generate password specific speaker models. Testing procedure involves the feature extraction and application of features to the models pertaining to recorded and original speech models. If the test speech belongs to the recorded speech, it is prevented from undergoing the further process. If it is an original speech, feature vectors of the test speech are applied to the password specific speaker models and based on the classification criteria, a speaker is identified and authenticated. Our system is found to be robust against playback attacks and has given better performance in authenticating sixteen speakers considered in our work. Passwords are isolated words and digits chosen from "TIMIT" speech database. This work is also extended to using AVSpoof database for authenticating 44 speakers against replay attacks and the performance is analyzed in terms of rejection rate.

Keywords Mel frequency perceptual linear predictive cepstrum (MFPLPC) · Probabilty · Playback attacks · Robustness · Speaker authentication · Vector quantization (VQ) · Replay attacks · Peak signal to noise ratio (PSNR) · Rejection rate

✉ C. Jeyalakshmi
lakshmikrce.2016@gmail.com

[1] Department of ECE/SEEE, SASTRA Deemed University, Thanjavur, India

[2] Department of ECE, K.Ramakrishnan College of Engineering, Samayapuram, Trichy, India

1 Introduction

The advancements in technology have made us to use biometrics to authenticate the persons with no need to have them physically. There are various human traits being used as biometric features which facilitate the provision for authorized persons only or restricting the unauthorized persons from making an intended service. Face, fingerprint, iris, retina and DNA are the most frequently used biometrics to authenticate the persons. Combination of these biometrics is also deployed to enhance the security in person authentication/verification. Specialized equipments and computers are necessary to extract these biometrics. Voice can also be used as a biometric in-person authentication. Cost effective microphones are only required in addition to the computers to record the speech. When voice is thought of as a biometric for speaker authentication, speaker recognition has evolved with two categories such as speaker identification and verification. Speaker identification is one - on - many mapping between the test speaker and all enrolled speaker models. Speaker verification is one-to-one mapping to check genuineness of the claim of the speaker for taking accept/reject decisions. Text-dependent speaker identification/verification involves the use of same speech utterance for training and testing and password specific person authentication falls into this category. Text-independent speaker identification/verification emphasizes the use of different speech utterance uttered by speakers for training and testing and person authentication can be done by using speech utterances uttered by the speakers and it is not necessary to remember the speech utterance spoken during the training session and speakers are not bound for any kind of fixed set of text utterances. Multilevel framework [3] has been done using speech as a biometric for person authentication. Voice is used as a biometric [24] for person authentication against replay attacks. Offline signature verification and speaker verification [22] system are combined as both of these biometric are accepted widely for person authentication. Speaker verification system [26] is implemented using whispered speech for voice-based biometric systems. Fingerprint and speech [4] are used as multimodal biometrics for mobile authentication applications. Face, fingerprint and speech [1] are used as multimodal biometrics for person authentication Face and speech are used as multimodal biometrics [5, 19, 25] for speaker verification. Speech-based biometric system [20] against voice conversion attacks is developed. Two dimensional PalmHash Code (2DPHC) [10–14] was used for secure palmprint verification. Techniques for action recognition from sensor- generated data are discussed [15] and novel approach for complex activity recognition is proposed. The feasibility of career path prediction from social network data has been scientifically and systematically studied [16]. An efficient algorithm to identify temporal patterns among actions and utilize the identified patterns to represent activities for automated recognition has been presented [17]. The water quality of a station [18] using a multitask multi-view learning method is introduced to fuse multiple datasets from different domains. Tracking generic human motion is proposed using a fusion formulation which integrates low and high dimensional tracking approaches into one framework [2].

In this work, playback audios are obtained by recording the speech samples present and saved at the same sampling rate of original speeches. The testing algorithm is developed to distinguish between recorded speech and original speech and authenticate the speakers using password specific speaker models. Feature extraction, recorded speech models, original speech models, password specific speaker models and testing procedure are appropriately developed to authenticate the speaker in a robust manner and the system has more immunity against playback spoofing attacks. The proposed algorithm is extended to check the authenticity of 44

speakers against replay–laptop and replay-laptop-HQ – speaker attacks by considering the speech utterances from AVSpoof database and the testing is done on genuine models.

2 Materials and methods

2.1 About the database

Speech database considered in this work contains speeches of 8 female speakers and 8 male speakers and speeches of all the speakers are taken for analysis to authenticate sixteen speakers. For each isolated word, there are 26 utterances uttered by the 16 speakers out of which 16 utterances of each speaker are used for training and the remaining 10 utterances are used for testing. For the classification between recorded speech and original speech, 256 utterances in all words are used for training and 160 utterances are used for testing. For speaker authentication based on original speeches, 16 password specific utterances are used for training and 10 utterances in each word are used for testing. Another database used in our work is AVSpoof database [6].

Data availability All relevant data are within the paper and its supporting information files.

2.2 Features based on cepstrum

The feature extraction algorithm stems from two ideas 1.Modeling of Vocal tract 2. Homomorphic filtering of speech. In the vocal tract model, speech has been produced by passing an excitation through a filter whose response models the effects of vocal tract on the excitation. Homomorphic filtering deals with the convolution of excitation and vocal tract response through the process of addition of logarithm of their transforms and vocal tract response have been separated out from excitation by linear filtering. Formant locations and bandwidth are used to indicate the variation between the speeches uttered by speakers. The features used for any classification must be reliable to perform any pattern recognition task. These features must have high inter-class variation and low intra-class variation among the classes. These features should also be representing the case study to improve the recognition/classification accuracy to enable the use of speech as a biometric for authenticating the speakers. MFPLPC speech analysis method [7–9, 23] gives the details about the perceptual features extraction with filters spaced in mel scale. The proposed method includes training phase and classification phase for distinguishing recorded speech and original speech initially and subsequently, classification is done to authenticate the speakers by applying features to all password specific speaker models.

2.3 Training phase

The overall flow diagram shown in Fig. 1 indicates the modules used for extracting the proposed features and creation of templates and testing procedure for authenticating sixteen persons with the speech utterances chosen from TIMIT database. Recorded speeches are created by recording the speeches already stored in a computer by using “Audacity” software and the recorded speeches are stored in a separate folder with same sampling rate as specified for original speeches. Sixteen recorded speeches of 16 speakers are concatenated before

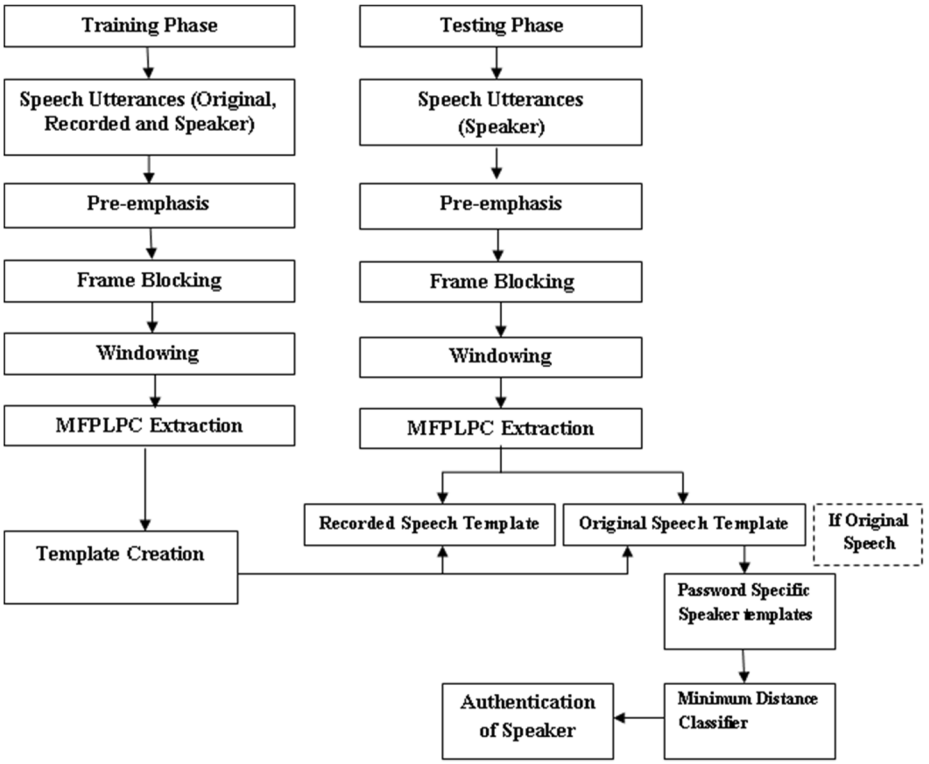

Fig. 1 The proposed person authentication system

extracting the features for recorded speech model creation. Similarly, original speech models are created by concatenating original speech utterances first, and the proposed features are extracted after performing the conventional preprocessing techniques such as pre-emphasis, frame blocking & windowing. Password specific speaker models are created in a similar manner.

2.4 Training model based on the clustering technique

Apart from reliable and promising feature extraction, one of the other important aspects is the generation of reference patterns or templates for pattern matching approach with respect to speech/language/speaker/emotion/noise recognition or classification. It is necessary to create templates as a representative model for individual speakers to achieve good accuracy for practical tasks so that speech can be used as a biometric for speaker authentication. For the given set of 'L" training vectors for recorded/original speech, VQ based clustering technique [21] is used as a pattern clustering approach to generate a set of "M" code books as representative templates for recorded and original speech. The basic idea of clustering approach is to reduce the signal's information rate through the use of codebook with relatively a small number of codewords as centroids as compared to the number of training vectors. In general, it is used to convert the training vectors into clusters and cluster centroids are representing the recorded/original speech and password specific speaker training data. The procedure [21, 23] used for converting the set of training vectors into a set of clusters is based on distance computation.

Classification procedure [21, 23] for arbitrary spectral analysis test vectors that chooses the codebook vector appropriate to the input vector and uses the code book vector as the resulting spectral representation. This is often referred to as nearest neighbourhood labeling or optimal encoding procedure. The classification procedure is essentially a quantizer that accepts the spectral analysis vectors as input and provides the output that is the codebook index of the codebook vector or the cluster centroids that best matches the input. It is done by computing Euclidean distance between each of the test vectors and M cluster centroids. The spectral distance measure for comparing features v_i and v_j is as in (1)

$$d(v_i, v_j) = d_{ij} = 0 \;\; when\, v_i = v_j \, and > 0 \;\; otherwise \tag{1}$$

If codebook vectors of an M-vector codebook are taken as y_m, $1 \le m \le M$ and new spectral vector to be classified is denoted as v, then the index m^* of the best codebook entry is as in (2)

$$m^* = \arg(\min(d(v, y_m))) for\, 1 \le m \le M \tag{2}$$

The formation of the clusters is done in such a way that the training data distribution is characterized and captured to produce a representative model or template for each speech/ speaker. It is analyzed that the most frequently occurring vectors do have less Euclidean distance as compared to the least frequently occurring ones.

2.5 Speaker authentication based on modeling technique

In general, any recognition system involves extraction of features from the training and test data, creating VQ codebook models for all the classes and feature vectors of each test utterance is tested against a certain number of template models to detect the identity of the class of that utterance from among the classes enrolled. Training data for recorded/original speech classification system is formed by concatenating the 160 speech utterances corresponding to the recorded/original speeches. Training data for speaker authentication is formed by concatenating the password specific speech utterances pertaining to the speaker. Other sets of 16 password specific speech utterances of a speaker are used for testing and 16 speakers comprising 8 male and 8 female speakers are considered for developing speaker authentication system.

Initially, features are extracted from the concatenated recorded or original speeches and the extracted features are applied to the training algorithm and reference models or templates are created for recorded/original speech. For feature extraction, recorded/original speech signal is passed through a pre-emphasis block, followed by frame blocking which converts the speech signal into overlapping frames of 16 msec duration for each block with 8 msec overlapping. Each frame is windowed by using a Hamming window [21] which is considered as the appropriate window for speech communication applications. Then the MFPLPC [7–9, 23] features are extracted. These feature vectors are applied to the VQ block to generate the codebook for recorded/original speech by adopting a K-means clustering procedure. In this algorithm, L training vectors are mapped with M clusters. Each feature vector in a block is normalized appropriately before giving as input to the module for generating reference templates for recorded/original speech. For testing, test data considered can be a recorded/ original speech utterance. The evaluation of the recorded/original speech classification system

is done by enabling the application of the perceptual features extracted from the test utterance to all templates corresponding to the recorded/original speech. Testing procedure initiates the process by first finding the minimum distance between each test vector and centroid of clusters. Average of minimum distances are determined for each speech model. The test utterance best associates with a recorded or original speech model which has a minimum of averages. If the test speech is classified as recorded speech, it is not allowed to participate in the process subsequently. For speaker authentication, feature vectors of the original test speeches are applied to the password specific speaker models and based on the minimum distance criterion, a speaker is authenticated or identified.

3 Results and discussion

The evaluation of the performance of the speaker authentication system using speech as a biometric based on the extraction of perceptual features on the recorded/original speech signal and clustering technique as a modeling technique is done by computing the squared Euclidean distance between feature vectors of the test utterance and template for recorded/original speech. If the test utterance is classified as recorded speech, it is prevented from undergoing a further process. If it is classified as original speech utterance, feature vectors of the original speech utterance are applied to the password specific speaker models and a speaker is authenticated based on the minimum distance classifier. The accuracy of the speaker authentication is the number of times the given speech utterances are correctly identified for a particular password speech utterance by the total number of test utterances corresponding to each speaker. The formula used for computing recognition accuracy is as in (3).

$$\%RA = \frac{Number\ \ of\ \ times\ \ the\ \ speaker\ \ is\ \ authenticated\ \ correctly}{Total\ \ln numbers\ \ of\ \ test\ \ utterances} \tag{3}$$

Figures 2 and 3 give the details about the description of the password “RUG BY” specific speech uttered by a female speaker and its recorded version in time and frequency domains.

The distance plot sown in Fig. 4 depicts the efficiency of the testing procedure used for discriminating recorded and original speech. Ten speeches are considered for testing. For each recorded speech, an average of minimum distances between test vectors and clusters of

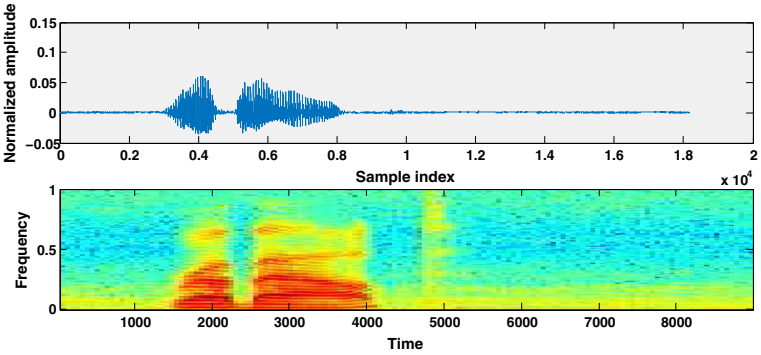

Fig. 2 Recorded speech “RUGBY” and its spectrogram

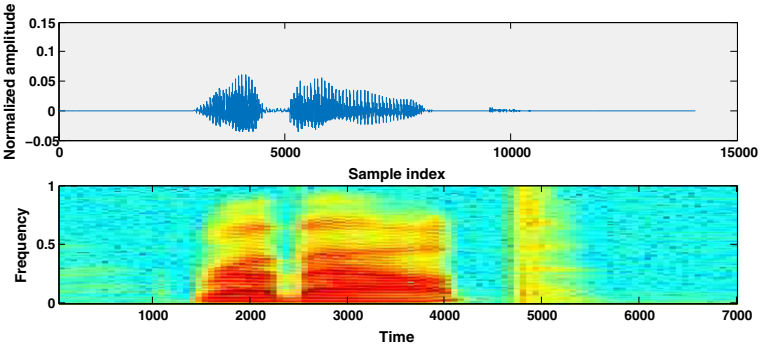

Fig. 3 Original speech "RUGBY" and its spectrogram

original training speeches is found to be minimum. If the test speech is classified as original speech, it will be permitted to undergo a further process of authenticating the speaker.

Figure 5 depicts the effectiveness of using a testing algorithm to isolate the recorded speech from the original speech. If the speech is classified as recorded speech by comparing the test vectors with recorded and original speech models, it will be ignored and stopped from further processing. For each recorded speech, the average of minimum distances is found to be minimum and this testing algorithm is more efficient in isolating the recorded test speech from the original speech i.e. providing immunity against playback attacks.

Figure 6 indicates the effectiveness in using the testing algorithm to discriminate speaker 2 from the enrolled ten speakers for each test speech by applying test vectors to all password specific speaker models. Passwords used are"enter, erase, go, help, no, rugby, repeat, stop, start, yes, zero, one, two, three, four and five" for eight female and eight male speakers considered in our work. Figure 7 depicts the discrimination of speaker 10 from the enrolled ten speaker models for each test speech.

Table 1 indicates the confusion matrix depicting the details about the speaker authentication by using voice passwords and it is providing 100% as accuracy for this small set of 16 speakers and the testing is done with ten test speeches for each speaker.

From the Table 1, it is evident that the algorithm developed is more robust in isolating recorded and original speeches and speaker authentication by applying the features of ten test

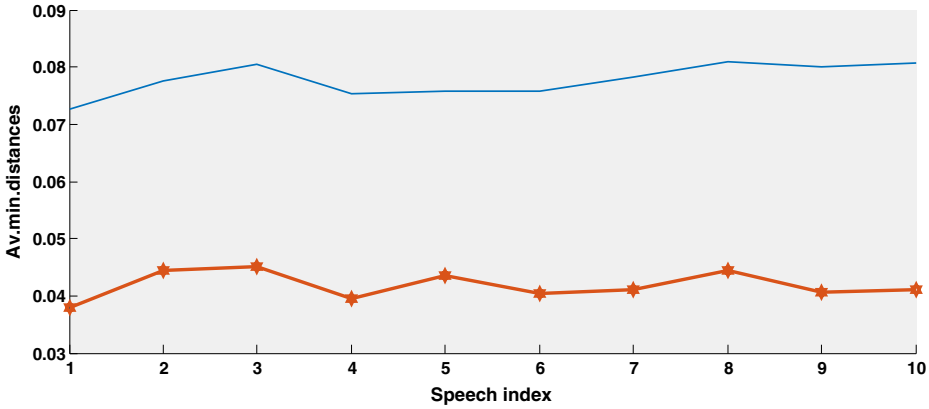

Fig. 4 Distance plot – discriminating original speech and recorded speech

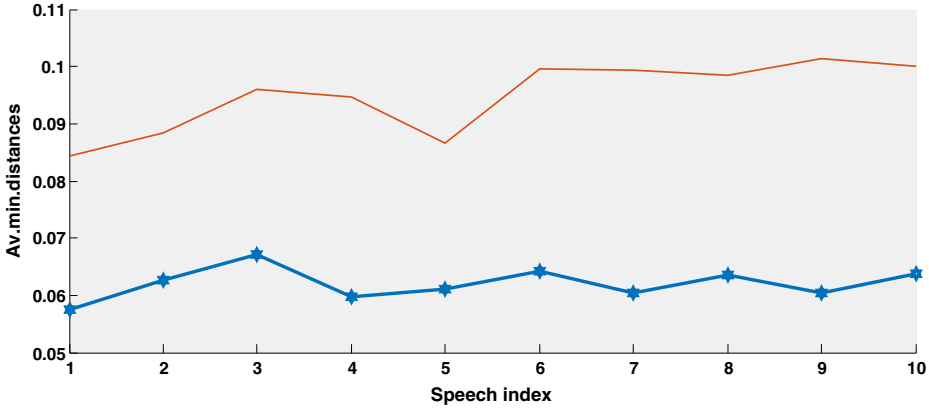

Fig. 5 Distance plot – isolation of the recorded speech from the original speech

speeches initially on models for discriminating original speeches from recorded speeches and subsequently, the features are applied to the password specific speaker models for authenticating the speaker from the set of sixteen enrolled speakers and the algorithm is said to be an immuniological approach against playback attacks. Performance of the system is also evaluated in terms of PSNR between original and recorded speech utterances of each speaker and is given in Table 2. High PSNR values indicate the closeness between original and recorded speech utterances. The proposed algorithm is more robust in discriminating recorded and original speech utterances by applying the features to the original and recorded speech models. If the speech utterance belongs to recorded speech model, it is forbidden from further processing. If the speech belongs to original speech model, it will further undergo the process to authenticate speakers.

This work is also extended for authenticating speakers against replay attacks using the speech utterances chosen from AVSpoof database [6]. Training phase involves the extraction of MFPLPC and MFPLPC concatenated with probability features from the "READ" genuine speech utterances and models/templates are created for 44 speakers comprising 31 male and 13 female speakers. Testing phase elucidates the extraction of features from the "PASS" genuine utterances and the features are applied to the models and based on the computation of the

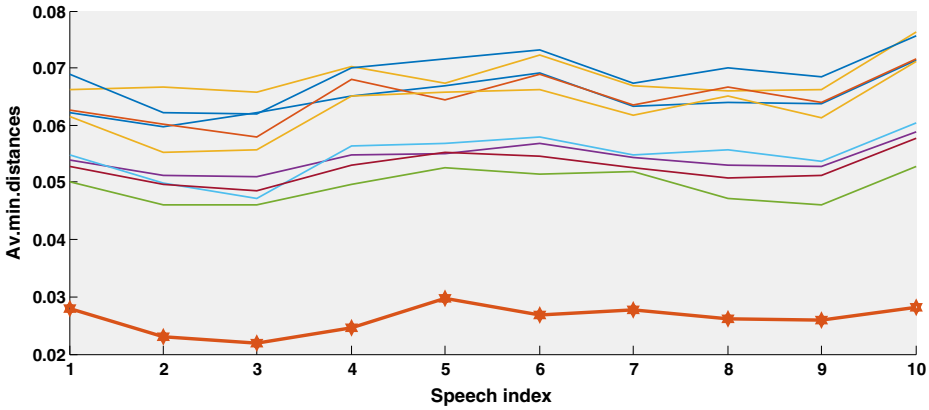

Fig. 6 Distance plot – authentication of speaker 2

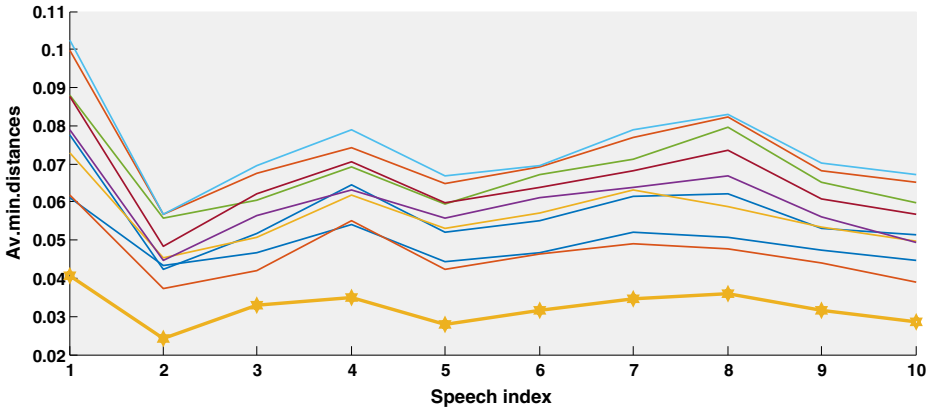

Fig. 7 Distance plot - authentication of speaker 10

distance between the test features and models, a speaker is authenticated. Authentication or the recognition accuracy (%RA) is calculated as, the number of times the speaker is authenticated correctly out of the total number of utterances considered for each speaker. Performance of the system is evaluated against the replay attacks such as Replay-laptop and Replay-laptop-HQ-speaker. Evaluation is done by extracting the features from the "PASS" attack utterances and applying these features to the genuine models. By using the minimum distance classifier, speaker classification is done. Figure 8 depicts the performance of the system against replay attacks for MFPLPC and MFPLPC concatenated with probability. Probability is computed for each frame as a ratio by finding the number of samples whose spectral energy is greater than the average spectral energy of the frame. The overall accuracy of the system is 88% for MFPLPC and 91% for MFPLPC concatenated with probability for testing the genuine test utterances against genuine models. The rejection rate is found to be 78 and 94.1% for the feature MFPLPC with the testing done on replay-laptop-HQ-speaker and replay-laptop atatck utterances against genuine models. Rejection rate obtained is 75.5 and 91.4% for the feature

Table 1 Confusion matrix – speaker authentication – MFPLPC Feature – TIMIT database

Train/ Test	Sp 1	Sp 2	Sp 3	Sp 4	Sp 5	Sp 6	Sp 7	Sp 8	Sp 9	Sp 10	Sp 11	Sp 12	Sp 13	Sp 14	Sp 15	Sp 16
Sp 1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sp 2	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sp 3	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
Sp 4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
Sp 5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0
Sp 6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
Sp 7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0
Sp 8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
Sp 9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0
Sp 10	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
Sp11	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
Sp12	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
Sp13	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0
Sp14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
Sp15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0
Sp16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10

MFPLPC concatenated with probability with the testing done on replay-laptop-HQ-speaker and replay-laptop attack utterances against genuine models.

Figure 9 indicates the efficiency of the feature MFPLPC concatenated with probability in

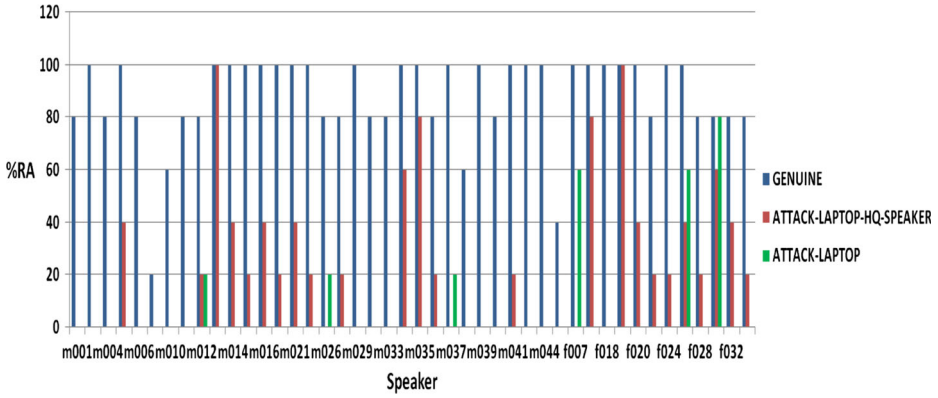

Fig. 8 Performance of the system against replay attacks – MFPLPC Feature – AVSpoof database

authenticating speakers against replay attacks.

The plots shown in Figs. 8 and 9 indicate the robustness of feature selection and modeling technique used to authenticate speakers against replay attacks with the speech utterances chosen from AVSpoof database.

4 Conclusions

This work discusses the application of perceptual features, perceptual feature fused with probability and clustering technique in building a robust speaker authentication system by using speech as a biometric against play back attacks with speech utterances chosen from

Table 2 Performance evaluation – PSNR between original and recorded speech

Speaker	Isolated word/digit as a password	PSNR between original and recorded speech utterances in dB
F1	RUGBY	41.4801
F2	REPEAT	42.0736
F3	STOP	42.7490
F4	START	42.2436
F5	YES	41.6877
M1	ENTER	45.5796
M2	ERASE	43.2415
M3	GO	41.6209
M4	HELP	42.8720
M5	NO	42.9070
F6	ZERO	39.6063
F7	ONE	33.4679
F8	TWO	36.4944
M6	THREE	41.0923
M7	FOUR	33.2370
M8	FIVE	33.2515

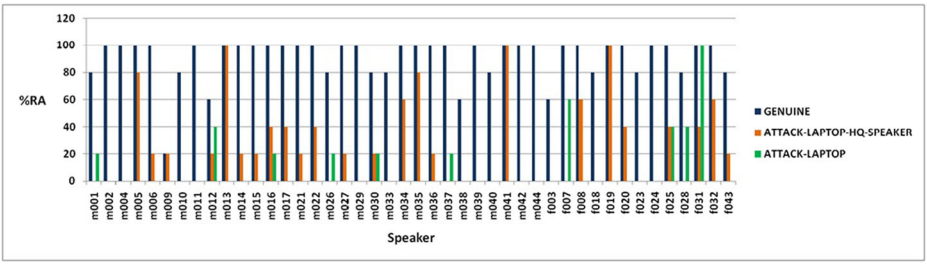

Fig. 9 Performance of the system against replay attacks – MFPLPC concatenated with probability AVSpoof database

TIMIT database. In this work, recorded speeches are considered as playback attacks. Speaker authentication system is developed by using the perceptual feature and VQ based clustering technique for recorded and original speech model creation initially. The testing procedure is used to discriminate the recorded and original speeches. Password specific training speeches for a speaker are concatenated for extracting features and subsequent model created for each speaker by using modeling technique. This work mainly emphasizes the importance and significance of modeling technique to isolate recorded speeches considered as playback attacks. Feature vectors of original speeches are applied to the password specific speaker models and based on the testing procedure, a speaker is authenticated. Performance of the system is evaluated by the number of times the speaker is authenticated correctly based on minimum distance classifier with respect to the password specific test speeches given for each speaker and this algorithm provides the maximum accuracy for this small set of speakers. This algorithm is highly robust to isolate the recorded speeches considered as playback attacks from original speeches and authenticate the speaker using voice as a biometric. This work is also extended to check the authenticity of 44 speakers against replay-laptop and replay-laptop-HQ-speaker attacks with the speech utterances chosen from AVSpoof database and performance of the system is evaluated in terms of rejection rate. The rejection rate is found to be very low for testing the genuine speech utterances on genuine models and is high for testing attack speech utterances on genuine models and it is revealed that the feature selection and modeling technique used are more robust against replay attacks also. The rejection rate is better for MFPLPC feature in comparison with MFPLPC concatenated with probability for replay-laptop-HQ and replay-laptop attacks.

Compliance with ethical standards

Competing Interest The authors have declared that no competing interest exists.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Bigun J, Fierrez-Aguilar J, Ortega-Garcia J, Gonzalez-Rodriguez J (2003) Multimodal Biometric Authentication using Quality Signals in Mobile Communications. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1234017&isnumber=27656
2. Cui J, Liu Y, Xu Y, Zhao H, Zha H (2013) Tracking generic human motion via fusion of low-and high-dimensional approaches. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(4). 996-1002. https://ieeexplore.ieee.org/document/6425496/

3. Das RK, Jeli S, and Mahadeva Prasanna SR (2016) Development of Multi-Level Speech based Person Authentication System. 1-13. https://springer.com/article/10.1007/s11265-016-1148-z
4. Dey S, Barman S, Bhukya RK, Das RK, Haris BC, Prasanna SRM, Sinha R (2015) Speech Biometric Based Attendance System. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6811345&isnumber=6811235
5. Duc B, Bigiin ES, Bigiin J, Maitre G, Fischer S (1997) Fusion of audio and video information for multi modal person authentication. 835-843. https://doi.org/10.1016/S0167-8655(97)00071-8
6. Ergünay SK, Khoury E, Lazaridis A, Marcel S (2015) On the vulnerability of speaker verification to realistic voice spoofing. Int Proc. Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS). https://ieeexplore.ieee.org/document/7358783/
7. Hermansky H, Morgan N (1994) RASTA processing of speech. IEEE transactions on speech and audio processing, 2(4): 578-589. https://www.ee.columbia.edu/~dpwe/papers/HermM94-rasta.pdf
8. Hermansky H, Tsuga K, Makino S, Wakita H (1986) Perceptually based processing in automatic speech recognition. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 11: 1971–1974. https://doi.org/10.1109/ICASSP.1986.1168649
9. Hermansky H, Margon N, Bayya A, Kohn P (1991) The challenge of Inverse E: The RASTA PLP method. Proceedings of Twenty Fifth IEEE Asilomar Conference on Signals, Systems and Computers 2:800–804. https://doi.org/10.1109/ACSSC.1991.186557
10. Leng L, Teoh ABJ (2015). Alignment-free row-co-occurrence cancelable palmprint Fuzzy Vault. International Journal on Pattern Recognition 48. 2290–2303. https://www.sciencedirect.com/science/article/pii/S0031320315000400
11. Leng L, Teoh ABJ 2017. Simplified 2D PalmHash code for Secure Palmprint Verification. International Journal of Multimedia Tools and Applications. 76(6). 8373-8398. https://link.springer.com/article/10.1007/s11042-016-3458-3
12. Leng L, Teoh ABJ, Li M, Khan MK (2014) Analysis of correlation of 2DPalmHash Code and orientation range suitable for transposition. international journal on Neurocomputing. 131. 377-387. https://www.sciencedirect.com/science/article/pii/S0925231213009351
13. Leng L, Teoh ABJ, Li M, Khan MK (2014) A remote cancelable palmprint authentication protocol based on multi-directional two-dimensional PalmPhasor fusion. Security And Communication Networks. 7. 1860–1871. https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.900
14. Leng L, Teoh ABJ, Li M, Khan MK (2015) Orientation range of transposition for vertical correlation suppression of 2DPalmPhasor Code. International Journalon Multimedia Tools and Applications 74(24): 11683–11701 https://link.springer.com/article/10.1007/s11042-014-2255-0
15. Liu Y, Nie L, Han L, Zhang L, Rosenblum DS (2015) Action2Activity: Recognizing Complex Activities from Sensor Data. Proceedings of the 24th International Conference on Artificial Intelligence. 1617-1623. https://arxiv.org/abs/1611.01872
16. Liu Y, Zhang L, Nie L, Yan Y, Rosenblum DS (2016) Fortune teller: Predicting your career path. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 201-207. https://dl.acm.org/citation.cfm?id=3015842
17. Liu Y, Nie L, Liu L, Rosenblum DS (2016) From action to activity: Sensor-based activity recognition. Iinternational journal on Neurocomputing. 181: 108–115. https://www.sciencedirect.com/science/article/pii/S0925231215016331
18. Liu Y, Zheng Y, Liang Y, Liu S, Rosenblum DS (2016) Urban Water Quality Prediction based on Multi-task Multi-view Learning. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16). 2576-2582. https://dl.acm.org/citation.cfm?id=3060981
19. McCool C, Marcel S, Hadid A, Pietikäinen M, Matějka P, Cernocký J, Poh N, Kittler J, Larcher A, Lévy C, Matrouf D, Bonastre J-F, Tresadern P, Cootes T (2012) Bi-Modal Person Recognition on a Mobile Phone: using mobile phone data. 635-638. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6266494&isnumber=6266221
20. Pal M, Saha G (2015) On robustness of speech based biometric systems against voice conversion attack. 30. 214-228. www.sciencedirect.com/science/article/pii/S1568494615000551
21. Rabiner L, Juang BH (1993) Fundamentals of speech recognition. Prentice Hall, New Jersey
22. Rani R, Sachdeva R (2016) Genetic Algorithm using Speech and Signature of Biometrics.03(12).240-245. https://www.irjet.net/archives/V3/i12/IRJET-V3I1299.pdf
23. Revathi A, Venkataramani Y (2011) Speaker Independent Continuous Speech and Isolated Digit Recognition using VQ and HMM. Proceedings of IEEE sponsored International conference on Communication and Signal processing:198–202. https://doi.org/10.1109/ICCSP.2011.5739300
24. Safavi S, Gan H, Mporas I, Sotudeh R (2016) Fraud Detection in Voice-based Identity Authentication Applications and Services.1074-1081. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7836786&isnumber=7836631

25. Sanderson C, Paliwal KK (2004) Identity verification using speech and face information. 449-480. https://doi.org/10.1016/j.dsp.2004.05.001
26. Sarria-Paja M, Senoussaoui M, Falk TH (2015) The effects of whispered speech on state-of-the-art voice based biometrics systems. 1254-1259. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7129458&isnumber=7129089

A. Revathi has obtained B.E (ECE), M.E (Communication Systems), and Ph.D (Speech Processing) from National Institute of Technology, Tiruchirappalli, Tamilnadu, India in 1988, 1993 and 2009 respectively. She has been serving on the faculty of Electronics and Communication Engineering for 30 years and she is currently working as a Professor in the Department of ECE, SASTRA Deemed University, Thanjavur, India. She has published 25 papers in Reputed International journals and presented papers in more than 30 International Conferences. Her areas of interest include Speech processing, Signal processing, Image processing, Biometrics & Security, Communication Systems, Embedded Systems and Computer Networks.

C. Jeyalakshmi received B.E degree in Electronics and Communication Engineering from Bharathidasan University in 2002 and M.E. degree in Communication systems from Anna University, Chennai in 2008. She served as a faculty for 11 years in the Department of ECE,Trichy Engineering college, Tamilnadu. Currently she has been with K.Ramakrishnan college of Engg., where she is working as Associate professor in ECE dept. She has obtained PhD degree from Anna University, Chennai in the field of Speech recognition of hearing impaired people in 2015. Her research interest also includes speech processing, Image processing, Neural networks, RF and Microwave Engineering.

K. Thenmozhi received her B.E (ECE) and M.E (Communication systems) degrees from Regional Engineering college (NIT), Tiruchirappalli and Ph.D. from SASTRA Deemed University, Thanjavur. Currently she is working as an Associate Dean in the Department of ECE at SASTRA Deemed University, Thanjavur. She has a teaching experience of 25 years. Her current research area includes secure Wireless communication and published 98 Research articles in International journals and conferences.