SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V10I10P112

Volume 10 Issue 10, 108-125, October 2023

Original Article

Integration of Timbrel, Cepstral Domain and Linear Prediction-Based Features for Replay Attack Detection

Amol A. Chaudhari[1], Dnyandeo K. Shedge[1], Vinayak K. Bairagi[1]

[1]*Department of E&TC Engineering, AISSMS Institute of Information Technology, Pune, India.*

[1]*Corresponding Author : amol191189@gmail.com*

Received: 12 August 2023 Revised: 14 September 2023 Accepted: 12 October 2023 Published: 31 October 2023

Abstract - *The automatic speaker verification system is vulnerable to several spoofing attacks. Among these spoofing attacks, detecting replay attacks is challenging as attackers do not need any expertise to mount replay attacks. Many efforts from the research community have focused on anti-spoofing solutions against the reply attack. Such efforts are classified as one focusing on feature extraction and others concentrating on classifiers. This work evaluates the performance of feature extraction schemes CQCC, LFCC, and MFCC. The success of Linear Prediction analysis has been demonstrated in the past. This work evaluates the performance of LPC and LPCC features. The recent work in the literature has focused on using multiple features and combining these features for improved performance. In this work, numerous components of CQCC, MFCC, LFCC, LPC and LPCC are integrated considering various combinations and evaluated. In literature, the success of Timbrel features has been demonstrated for speaker identification. The feature vector formed using various Timbrel features is integrated with cepstral and linear prediction-based features. Finally, Timbrel features zero cross rate are combined with these multiple features. Among all experiments carried out on the ASVspoof 2017 version 2 database, EER 5.44% is achieved for the integration of zero cross rate and LPC on the development set and 17.79% EER is conducted for the integration of zero cross rate, MFCC, CQCC, LFCC, and LPCC features on evaluation set.*

Keywords - *Automatic Speaker Verification, Replay attack, Timbrel features, T-SNE, Zero cross rate.*

1. Introduction

A biometric system intends to verify a person's biological and behavioural characteristics [1]. The classification of body traits that can be used for biometric recognition includes anatomical and behavioural features [1, 2] Anatomical characteristics listed in [1] include iris, face, hand geometry, fingerprint, palm print and ear shape. In contrast, some behavioural features are signature, gait, and keystroke dynamics [1]. "Voice biometrics can be considered either as an anatomical or as a behavioural characteristic" [1, 2]. It is essential to have a robust and secure system from the deployment perspective. Speaker identification and speaker verification are part of speaker recognition [3]. Speaker identification is the system which identifies who the speaker is, and speaker verification is the system which verifies the claimed identity, whether true or false. Applications of Automatic Speaker Verification (ASV) systems include access to systems that handle classified information and banking transactions [4]. Only enrolled speakers and users can access the ASV system [4]. Non-genuine or imposter speakers are those who do not have access. However, the imposter/attacker deliberately attempts to gain unauthorized access to the ASV system, called spoofing attacks on ASV [4].

The spoofed speech samples can be produced through speech synthesis, voice conversion, or the replay of recorded speech. Spoofing attacks can be divided into direct and indirect attacks based on how the spoof samples are presented to the ASV system. Through the sensor, the samples are used as input for the ASV system in direct attacks (also known as Physical Access (PA) attacks); direct attack is at the transmission and microphone level [1]. The samples passing the sensor, or the ASV system software process, are subject to indirect attacks, often called Logical Access (LA) attacks, which entail accessing the samples during feature extraction, tampering with the models, and decision-making or score computation stages [1]. Figure 1 shows the block diagram of a spoofing attack considering direct and indirect attack points. Automatic Speaker Verification (ASV) system is vulnerable to various types of spoofing attacks, including speech conversion, impersonation text-to-speech synthesis and replay attacks. The research community focused on robust countermeasures for detection of spoofed speech. Several challenges were held in the past for developing novel approaches for detecting spoofed speech. Table 1 summarizes the challenges faced during INTERSPEECH workshops.

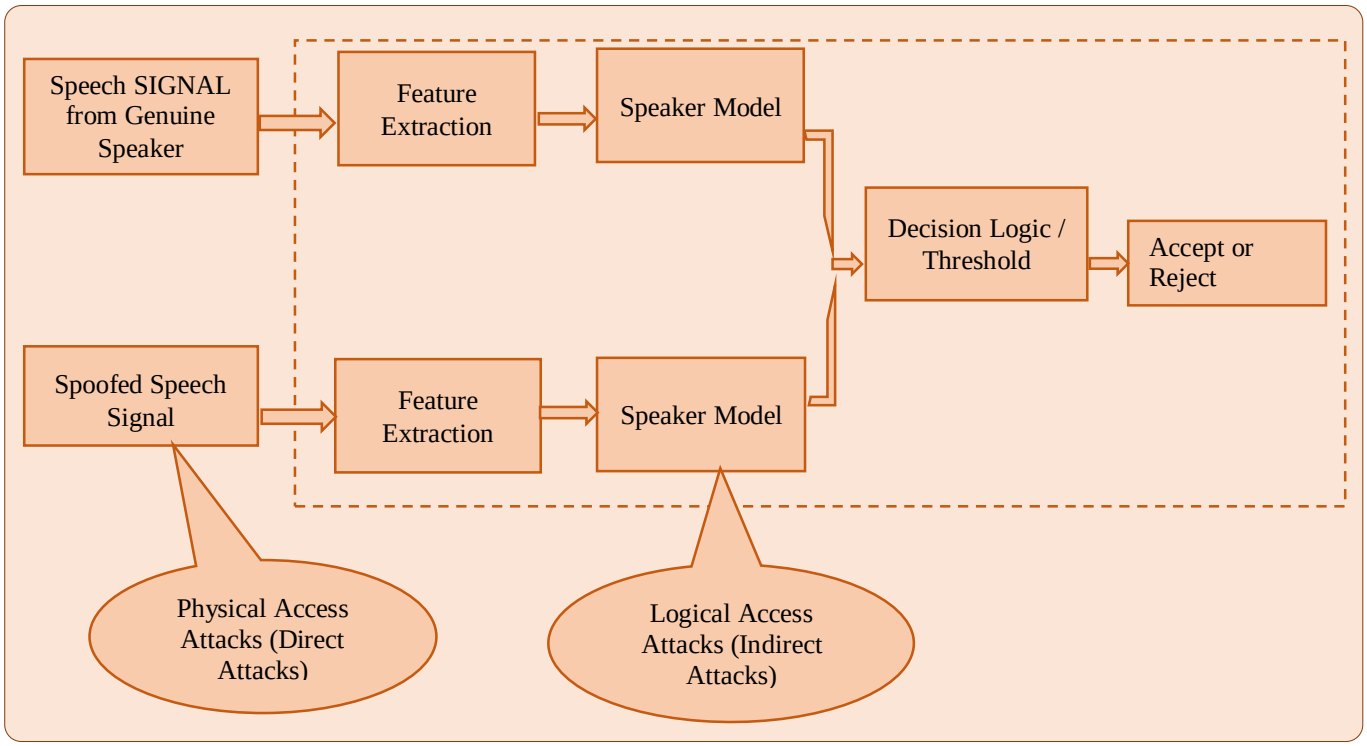

Fig. 1 Block diagram of spoofing attack considering direct attack and indirect attack points [1]

Table 1. Summary of ASVspoof challenges

Sr. No.	Challenge	Comment
1	INTERSPEECH 2013 [5]	Details of different ASV system vulnerabilities and their corresponding mitigation strategies were published.
2	ASVspoof 2015 Challenge [6]	Concentrated on developing several countermeasures against speech synthesis and voice conversion spoofs
3	ASVspoof 2017 challenge [7]	The focus was exclusively on replay spoof speech detection
4	ASVspoof 2019 challenge [8]	The emphasis was on synthetic or simulated replay
5	ASVspoof 2021 challenge [9]	Three-track challenge including LA, PA, and DeepFake detection

Among these attacks, replay attacks are most accessible for the imposter. In a replay attack, the target speaker's recorded voice is used to get unauthorized access. Detecting replay attacks is more challenging because of high-quality recording devices and playback systems. Also, the attacker/imposter requires no skills or technical knowledge to mount the replay attack. ASVspoof 2017 challenge mainly focused on developing countermeasures for detecting replayed spoofed speech. There have been several efforts to detect replay attacks in the ASVspoof 2017 challenge. The ASVspoof 2017 challenge focused on developing countermeasures using acoustic characteristics of genuine versus replayed speech [10]. The literature analysis for the work carried out by researchers in the ASVspoof 2017 challenge is discussed next. In [11], CQCC, MFCC, LFCC, IMFCC, RFCC, LPCC, SCFC, SCMC, and SSFC are compared for replay attack detection. Out of these features, it is mentioned in [11] that Subband Spectral Centroid Magnitude Coefficients (SCMCs) perform better for replay attack detection. Another approach proposed for the replay attack detection includes VESA-IFCC features [12], score level fusion of IFCC, CQCC and MFCC [13], high frequency analysis of IMFCC, LPCC, LPCCres, CQCC, and Cepstrum features [14].

The work in [15] systematically analyzed the state-of-the-art voice Presentation Attack / Spoofing Attack Detection (PAD) systems. As mentioned in [15], the work carried out by various researchers may use fusion or not. Two main fusion techniques score fusion and feature fusion were applied in the works that employ fusion. Other fusion techniques are available; however, they are few. To consider multiple scores produced by voice PAD models in the classification decision, score fusion is used [15-17].

Table 2. Summary of approaches from literature focused on using multiple features and fusion

Sr. No.	Author	Features	Database	EER (%)
1	Gupta et al., 2023 [4]	Score Level Fusion of CQCC, CFCC, CFCCIF, CFCCIF-ESA, CFCCIF-QESA	ASVspoof 2017 Version 2	9.21% EER on the Development Set and 11.24% EER on the Evaluation Set
2	Dutta et al. 2021 [22]	SCQCC+GMFCC	ASVspoof 2017 Version 2	8.60% EER on Evaluation Set
3	Kamble and Patil 2021, [10]	CQCC+LFCC+MFCC +TECC	ASVspoof 2017 Version 2	6.68% EER on the Development Set and 10.45% EER on the Evaluation Set
4	L. Liu and Yang, 2020 [23]	Concatenated Features: CQEPIC	ASVspoof 2019	EER 6.97%
5	Kamble, Tak, et al., 2020, [24]	AM-FM Demodulation Based Features	ASVspoof 2017 Version 2.0	Reduction in EER: 11.93% for AM and 10.12% for FM-Based Features
6	Balamurali et al., [25]	MFCCs, Spectrogram, CQCCs, LPCCs, IMFCCs, RFCCs, LFCCs, SCFCs, SCMCs, and CCCs	ASVspoof 2017	EER 10.8%
7	Phapatanaburi et al., [26]	Combination of Linear Predication Analysis, Residual Phase and CQCC	ASVspoof 2017 Version 2	EER 9.26 %
8	Singh & Pati, 2019, [27]	RMFCC + CQCC	ASVspoof 2017	EER 9.50%
9	Singh & Pati, 2019, [28]	Score Level Fusion of LPRHEMFCC, RPCC and CQCC	ASVspoof 2017	EER 8.86%
10	Oo et al., [29]	Combination of CQCC and Gammatone-Scale RP	ASVspoof 2017	EER 9.48% on the Evaluation Set

To combine these scores, many methods are employed, including the mean, sum, standard deviation, min, max, weighted or normalized sum, etc. [15] Feature fusion approaches are also used in the Multimodal Biometric Identification system. Feature fusion is done either serial or parallel to boost the recognition rate. [15, 18]. Various feature vector sets are serially combined into a single feature vector through serial fusion. Parallel feature fusion is based on a complex vector, a vector with components of complex numbers, as opposed to serial fusion, which is based on the union vector.

A thorough literature analysis of fusion-based approaches for detecting replay attacks is discussed. Better discrimination of genuine voice from spoof voice can be achieved using complementary information present in multiple features [15, 20-22]. Hence, most of the work used numerous features [15]. Table 2 summarises approaches that focus on using multiple features and fusion.Most of the work in the literature focussed on using multiple features and fusion of elements for improved performance against replay attack detection. However, the effectiveness of Timbrel features needs to be evaluated for replay attack detection. Also, the fusion of Timbrel features with the cepstral domain and linear prediction-based features needs to be assessed. This work analyses the performance of features such as LPC, LPCC, and LFCC. The rigorous analysis is carried out by integrating CQCC, MFCC, LPC, and LFCC features, considering several combinations. Energy in residual is also integrated with LPC and LPCC.

Table 3. Details of ASVspoof 2017 dataset version 2

Database Subset	Number of Speakers	Genuine Utterances	Spoofed Utterances
Training	10	1507	1507
Development	8	760	950
Evaluation	24	1298	12008

Most often, zero-crossings are used as audio features in speech-processing applications the use of zero crossings as audio features is demonstrated in [30] for speech recognition. The success of Timbrel features for recognizing speakers for whispering speech has been shown in [31-33]. In [31], zero cross rate is used as a timbre feature for speaker identification of whispering speech. This work analyses Timbrel features by integrating them with cepstral features. Also, zero-crossings are combined with cepstral features CQCC, MFCC, LFCC, and LPC and LPCC. The performance of all components is evaluated on the ASVspoof 2017 version 2 development and evaluation set. A detailed comparison of several integrated features is carried out. This paper is organized as follows. After the introduction in Section 1, the proposed methodology is presented in Section 2. In section 3, experimentation is discussed, followed by results in section 4. Section 5 offers the overall work carried out in discussion. This work is concluded in section 6.

2. Proposed Methodology

2.1. Database

The experimentation has been carried out on the ASVspoof 2017 version 2 database [34, 35]. The genuine utterances in this database are from the RedDots corpus [34]. As mentioned in [34], natural statements are replayed and recorded using various diverse devices and acoustic environments for obtaining spoofed utterances. Training, development and evaluation are the three non-overlapping subsets present in this database. More details on this database can be referred to from [34]. Table 3 presents the statistics about the ASVspoof 2017 version 2.0 database.

2.2. System Design

The block diagram of the proposed system is shown in Figure 2. The system design follows the MATLAB-based reference given by ASVspoof 2017 organizers [35]. In the training phase, features are extracted from genuine and spoof data using the ASVspoof 2017 version 2 database train data. In the proposed methodology, different components are concatenated after the feature extraction step, and the feature set cell is given as input to GMM. In the next section, the various elements used for analysis are discussed. Two different GMMs are trained, one on the training dataset's genuine voice utterances and the other on spoofs. In the development phase, features are extracted from development data. Given the natural and spoofed speech models, the score is calculated as the log-likelihood ratio for the test utterance. The algorithm of the proposed work is shown in Figure 2, and Figure 3 presents the algorithm for implementing the methodology presented in Figure 2.

2.2.1. Pre-Processing

Pre-processing of speech signals involves pre-emphasis, framing, and windowing. The higher frequencies diminished during speech production are boosted using pre-emphasis followed by framing. As speech signal is assumed to remain stationary for 20-30 milliseconds, the pre-emphasized signal is segmented into short segments of 20-30 millisecond frames with ten milliseconds overlap. In the next step, the windowing of each speech frame is carried out. A hamming window is generally preferred.

2.2.2. Feature Extraction Approaches for Replay Attack Detection

CQCC

The CQCC feature is more suited to the ASVspoof task since it possesses variable spectrum resolution and efficient time-frequency representation to identify replayed voices. [29]. As a benchmark feature that has been successfully used in the ASVspoof system, the CQCC feature is used [29]. The constant-Q transform of input discrete time-domain signal x(n)is calculated as,

$$X^{CQ}(k,n) = \sum_{j=n-\frac{N_k}{2}}^{n+\frac{N_k}{2}} x(j) a_k^* \left(j - n + \frac{N_k}{2}\right) \quad (1)$$

Where,

k = 1, 2, . . ., K is the frequency bin index,
a_k^* is the complex conjugate of a_k, and
N_k are variable window lengths

Geometrically spaced bins ensure a constant Q factor. It is a ratio of the centre frequency to the bandwidth [36]. A higher frequency resolution at lower frequencies and higher temporal resolution at higher frequencies is provided by CQT [36]. Next step, the power spectrum of $X^{CQ}(k,n)$ is computed, then followed by a log. Then, using uniform re-sampling, the geometric space is transformed into a linear space for cepstral analysis. [29]. Last, DCT is computed to get the CQCCs. The following equation represents CQCC computation.

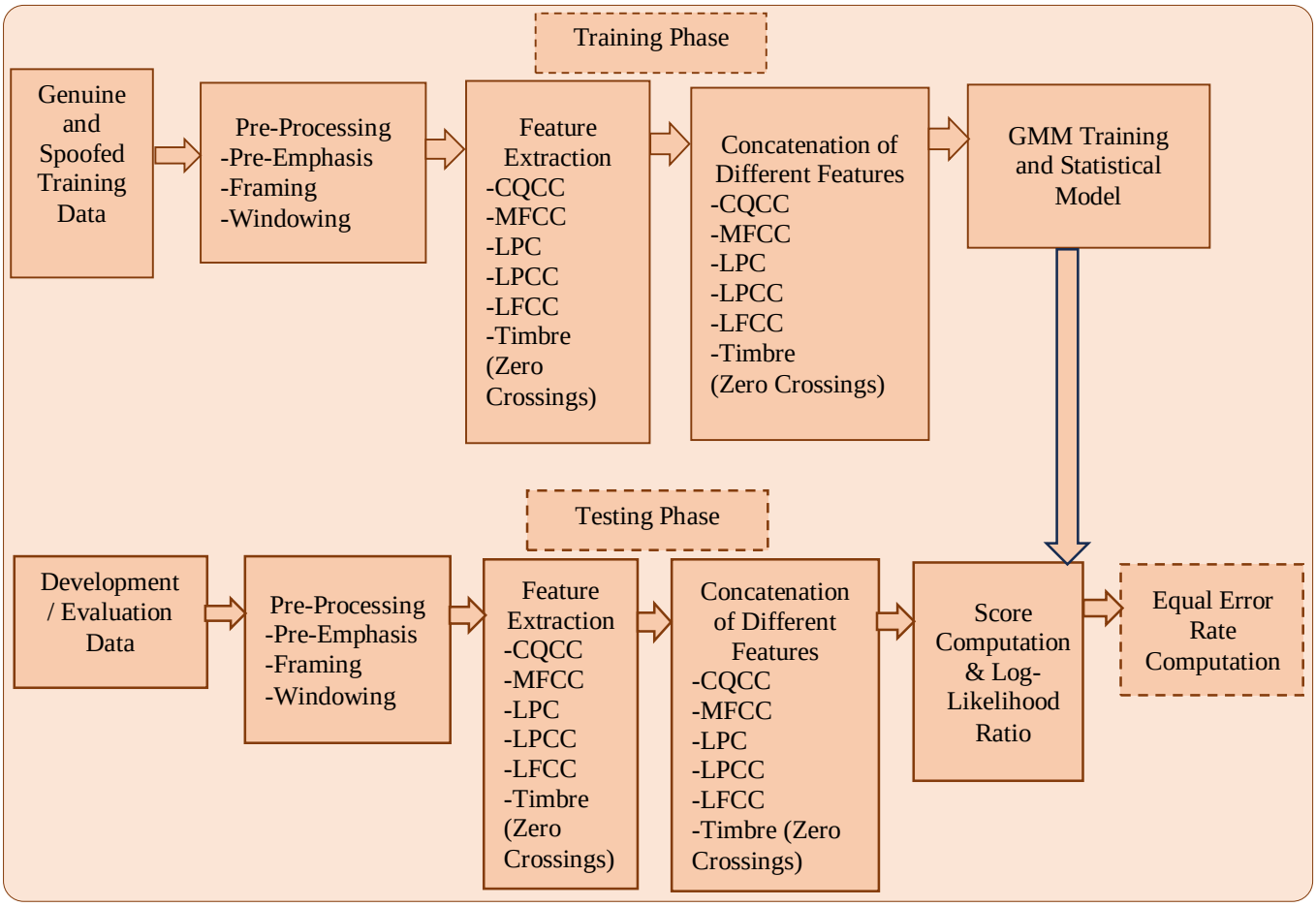

Fig. 2 Proposed block diagram of speaker recognition system

Algorithm for Proposed Work:

```
For N = genuineID from the training set
        Read each speech signal from the database path
        Pre-processing of genuine speech signal
        Extract different speech features
        Integrate extracted features
        Train GMM for genuine samples from extracted and integrated features.
end
For N = spoofID from the training set
        Read each speech signal from the database path
        Pre-processing of spoof speech signal
        Extract different speech features
        Integrate extracted features
        Train GMM for spoof samples from extracted and integrated features
end
For N = genuineID or spoofID from the development or evaluation set
        Read each speech signal from the database path
        Pre-processing of speech signal
        Extract different speech features
        Integrate extracted features
        Compute scores and log-likelihood ratio
        Compute Equal Error Rate
end
```

Fig. 3 Algorithm of the proposed methodology

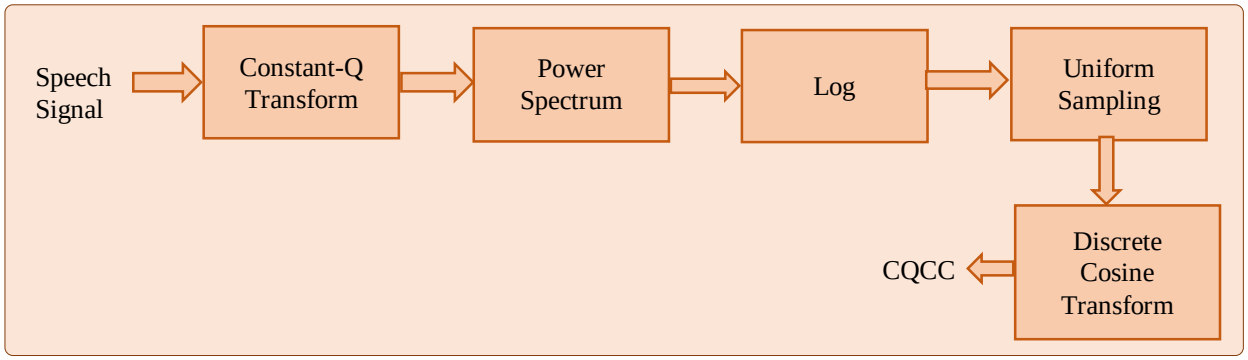

Fig. 4 Block diagram of CQCC feature extraction

$$CQCC(r) = \sum_{l=1}^{L} log\,|X^{CQ}(l)|^2 \cos[\frac{r(l-\frac{1}{2})\pi}{L}] \quad (2)$$

Where,

r = 0, 1, . . ., L -1, and

l is the newly re-sampled frequency bins.

From [36], details on CQCC feature extraction can be referred. Figure 4 shows the block diagram of CQCC feature extraction.

MFCC

In speech processing, one of the well-known magnitude-based features is the MFCC [29, 37]. MFCC is based on the cepstral analysis using log magnitude spectrum on a mel scale. Figure 5 shows the block diagram of MFCC. In the computation of MFCC, framing of speech signal is performed. Generally, the speech signal is divided into 20-30 milliseconds frames with 25-50% overlap. Next, the framed speech signal's windowing (usually hamming) is computed, followed by a fast Fourier transform. Using a fast Fourier transform, the power spectrum is calculated. The filter bank processing is then done using mel-scale on the power spectrum. The power spectrum is translated into a log domain followed by DCT to obtain MFCCs. The following equation [38] represents the calculation of MFCCs.

$$\widehat{C_n} = \sum_{n=1}^{k} (log\ \widehat{S_k})\ \cos[n(k-\frac{1}{2})\frac{\pi}{k}] \quad (3)$$

Where k is the number of Mel cepstrum coefficients, $\widehat{S_k}$ is the output of the filterbank and $\widehat{C_n}$ is the final mfcc coefficient.

LPC and LPCC

By minimizing the mean square error between the input speech and estimated speech, the linear prediction approach is used to obtain filter coefficients corresponding to the vocal tract [38]. LPC represents a current speech as a linear combination of previous samples. The following equation in [32] describes the LPC calculation.

$$x(n) = \sum_{k=1}^{p} a_k\, x(n-k) \quad (4)$$

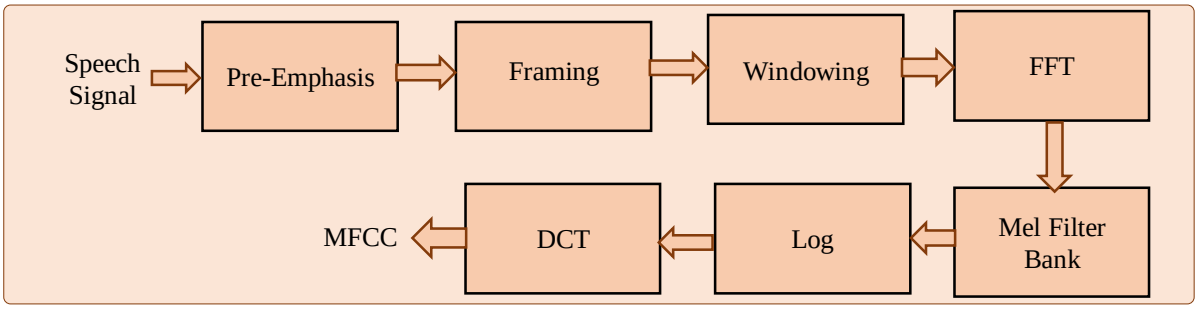

Fig. 5 MFCC feature extraction

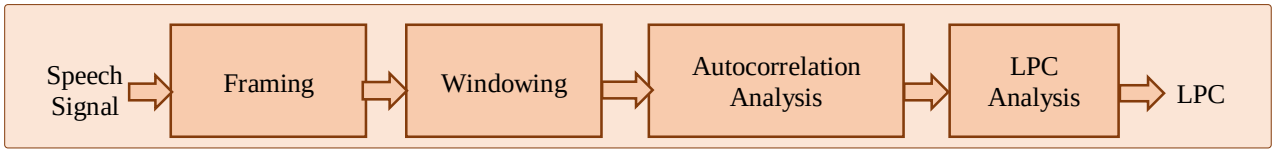

Fig. 6 LPC feature extraction [38]

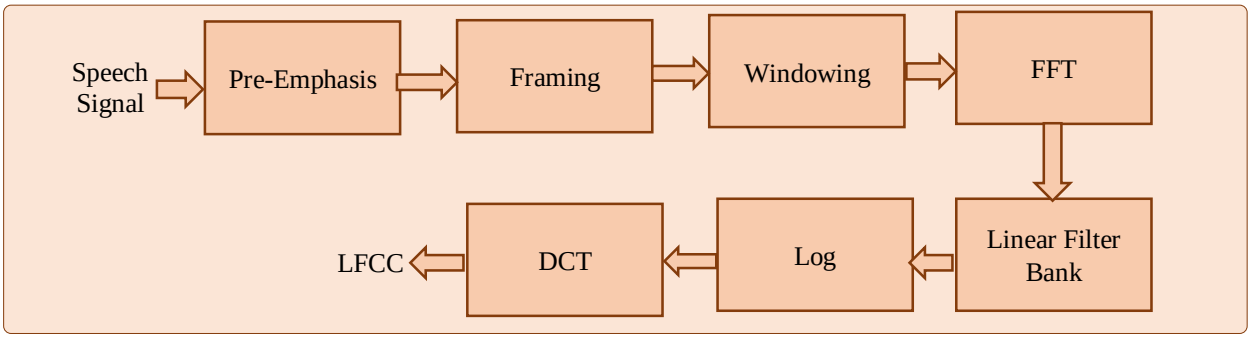

Fig. 7 LFCC feature extraction

Where,

The predicted signal is x(n),
The previous sample is x(n - k),
The predictor constant is a_k.

Figure 6 presents the block diagram of LPC. The energy residual is obtained in the computation process of LPC. LPC-calculated spectral envelope is used for the computation of cepstral coefficients, namely, Linear Predictive Cepstral Coefficients (LPCC) [38, 39].

LFCC

Like MFCCs, linearly spaced triangular filters extract Linear Frequency Cepstral Coefficients (LFCCs) [40]. Figure 7 presents a block diagram of LFCC feature extraction. First, the integration of the power spectrum with overlapping band-pass filters is carried out. After this integration, logarithmic compression and DCT are performed to compute the cepstral coefficients.

Timbrel Features

In speech processing applications such as speech recognition, zero crossing rate is often used as an audio feature. An audio frame's Zero-Crossing Rate (ZCR) measures how frequently the signal's sign shifts across the frame. It is calculated as the number of times the signal's value switches from positive to negative and vice versa, divided by the frame's length. Zero crossing-based feature extraction has been demonstrated in [30] for speech recognition. Another approach mentioned in [31-33] has proposed using Timbrel features for speaker recognition. Even when two sounds are presented similarly, a listener can distinguish between nonidentical ones because of the multidimensional and perceptual quality of the timbre. [32, 41]. One of the Timbrel feature zero-cross rates is used for speaker identification of whispering speech [31]. This work uses zero cross-rate Timbrel features to test the effectiveness of detecting replay attacks. Also, zero cross rate is integrated with cepstral features (CQCC, MFCC, LFCC, LPCC) and LPC.

2.2.3. Classifier

The most often used Classifier is still GMM [15, 22]. As a baseline system using a CQCC-based feature, the ASVspoof 2017 challenge also offered the GMM classifier. As a result, in this work, GMM classier is employed to distinguish between genuine and replay voice samples. As mentioned in the ASVspoof 2017 challenge, two separate GMMs are trained on the training dataset's simple and spoofed speech utterances, respectively, with 512-component models trained with an Expectation-Maximization (EM) algorithm with random initialization. Given the natural and spoofed speech models, the score is calculated as the log-likelihood ratio for the test utterance.

2.2.4. Evaluation Metric

ASVspoof 2017 challenge has provided an evaluation metric as an Equal Error Rate (EER) for baseline systems using the Bosaris toolkit [42]. When assessing the effectiveness of replay detection systems, EER is the primary metric employed [22]. The rate at which the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal is known as the Equal Error Rate (EER) (decision threshold). The details on FAR and FRR are in [22]. LLR scores are used to compute FAR and FRR. The Receiver Operating Characteristics Convex Hull (ROCCH) is used to calculate the EER in the Bosaris_toolkit, which is used to estimate EERs [22]. The percentage EER is calculated as mentioned in the following equation,

$$\% \, EER = \frac{FAR+FRR}{2} \times 100 \, \% \qquad (5)$$

3. Experimentation

To analyze genuine and spoof voices, natural speech samples with id T_1000003 and spoof speech sample with id T_1001511 is considered from the ASVspoof 2017 version 2 database. These samples are selected from the training set of the database. The genuine and spoofed speech signal is analyzed using PRAAT [43] software. Figure 8 and 9 shows natural and spoofed speech signal, respectively. From figures 8 and 9, it is observed that genuine speech signal is relatively smooth and periodic as compared to spoofed signal. Also, a spoofed signal has a relatively high amplitude with noise compared to a simple signal. This differentiation can be because recording device properties are added to the original speech signal.

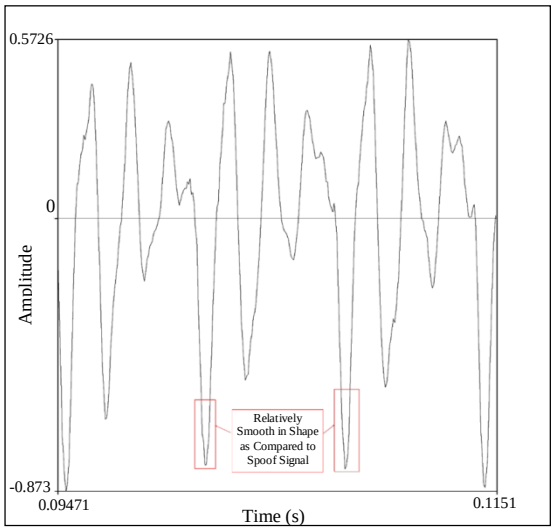

Fig. 8 Genuine speech signal

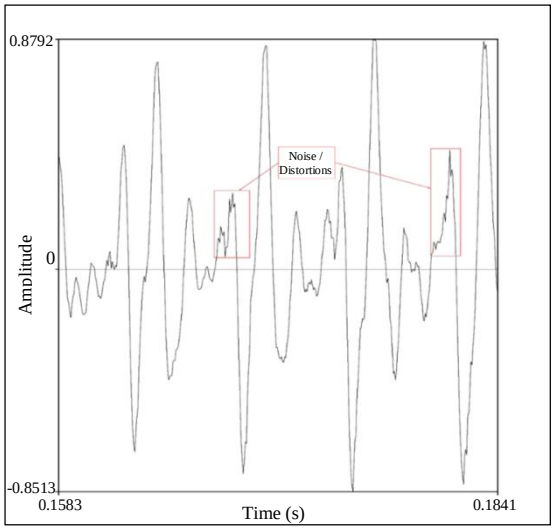

Fig. 9 Spoof speech signal

3.1. CQCC and MFCC Feature Extraction/Computation

The current authors have evaluated the baseline CQCC, MFCC features and integration of CQCC and MFCC features on the ASVspoof 2017 version 2 development set the work presented in [44]. This work compares the results in [44] with the different feature extraction schemes and concatenation approaches. More information on CQCC and MFCC parameters used in this work is referred from [44]. CQCC and MFCC feature extraction techniques calculate the features in the form of high-dimensional data, i.e., number of rows * number of columns. These dimensions differ depending on factors such as the length of the speech signal and the nature of the speaker, whether they are uttering fast or slow. However, the number of rows representing the number of filters or coefficients is kept constant for all speech samples of the database. For example, the dimension of CQCC features for model T1000003 is 90*177.

High-dimension data can be visualized using techniques such as T-SNE [40]. T-SNE analysis is carried out with ninety-dimensional (including delta and double delta) CQCC reduced to two -dimensional feature vector. At the same time, feature vectors of MFCC are considered with eighty-four dimensions (including delta and double delta) reduced to a two-dimensional feature vector. MFCC's eighty-four-dimensional feature vector is considered by eliminating the vectors calculated to zero-valued. This elimination is only viewed for data visualization the Figure 10 and 11 T-SNE representation of CQCC and MFCC features, respectively.

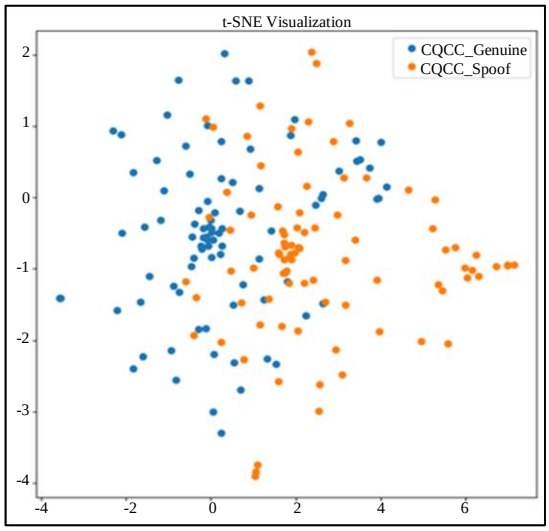

Fig. 10 CQCC feature distribution for genuine and spoof sample

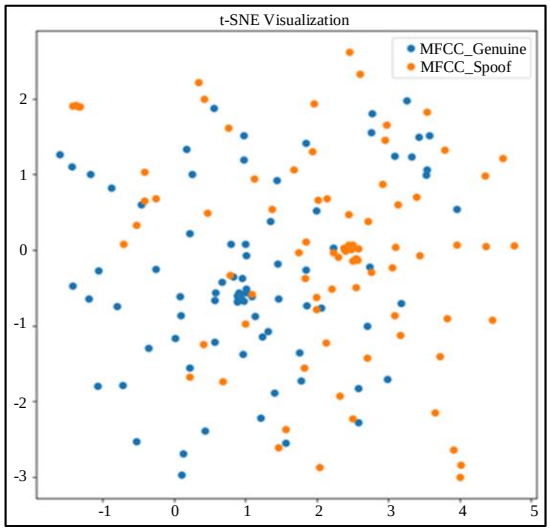

Fig. 11 MFCC feature distribution for genuine and spoof sample

Figures 10 and 11 show that genuine speech samples of MFCC are overlapped with the spoofed speech samples. Whereas, in the case of CQCC, this overlap is reduced compared to MFCCs.

Table 4. LPC parameters

Parameter	Statistics
Frame Length	512 samples
Overlap	50 samples
Window	Hamming
Number of Coefficients	12

3.2. LPC and LPCC Feature Computation

The LPC coefficients computed in [44] have not considered/mentioned pre-processing steps framing, overlapping, and windowing for speech signal. This work involves prior computation of LPC coefficients, framing, overlapping, and windowing of the speech signal.

Table 4 shows the parameters used in the extraction of LPC coefficients. "lpcauto" function from the voicebox toolbox [45] is used to compute LPCs based on the autocorrelation method. In this work, the first coefficient, "1", is excluded/omitted from the LPC feature vector in all experimentations.

"lpcauto" function computes energy in the residual. In this work, power in the residual is also analyzed for replay attack detection. Several sub-routines are provided in the voicebox toolbox for converting LPC coefficients to various forms, including complex coefficients. "lpcar2cc" sub-routine is used to extract LPCCs from LPC coefficients. LPC feature vector with first coefficient "1" is used to compute LPCC in all experimentations.

Twelve LPCCs are considered for performance evaluation. For data visualization in the case of LPC, a ninety-dimensional feature vector is converted into a two-dimensional vector.

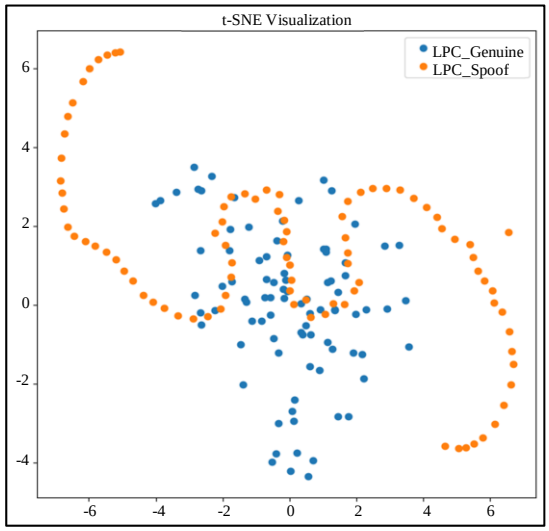

Fig. 12 LPC feature distribution for genuine and spoof sample

Figure 12 shows the T-SNE visualization of LPC for genuine and spoofed speech signals. T-SNE representation of LPC for spoofed speech samples differs from that observed in all cases. This indicates that LPC features are more effective in non-overlapping between genuine and spoof signals.

3.3. LFCC Feature Computation

ASVspoof 2019 organizers provided a baseline system, which is evaluated. The LFCC parameters used are the same as those offered by ASVspoof 2019 organizers. Table 5 shows the parameters used in the LFCC feature extraction.

The number of filters was considered as 30, which provides 30 LFCC coefficients, 30 delta and 30 double delta coefficients. LFCC features are concatenated with several features mentioned in this work. For data visualization in the case of LFCC, a ninety-dimensional feature vector is converted into a two-dimensional vector.

The Figure 13 shows the T-SNE visualization of LFCC for genuine and spoofed speech signals. Figure 13 shows that the LFCC features of genuine speech and spoofed speech show significant non-overlap compared to CQCC and MFCC.

Table 5. LFCC parameters

Parameter	Statistics
Sampling Frequency in Hz	16000
Window Length in milliseconds	20
Number of FFT bins	512
Number of Filters	30
Number of Coefficients	30 Including 0^{th} Coefficient + Δ + ΔΔ

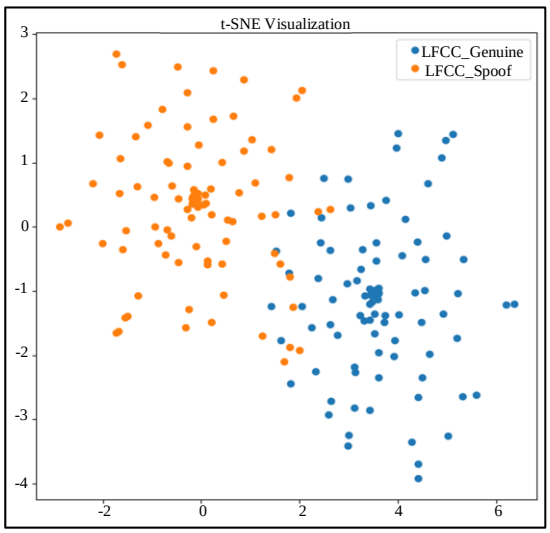

Fig. 13 LFCC feature distribution for genuine and spoof sample

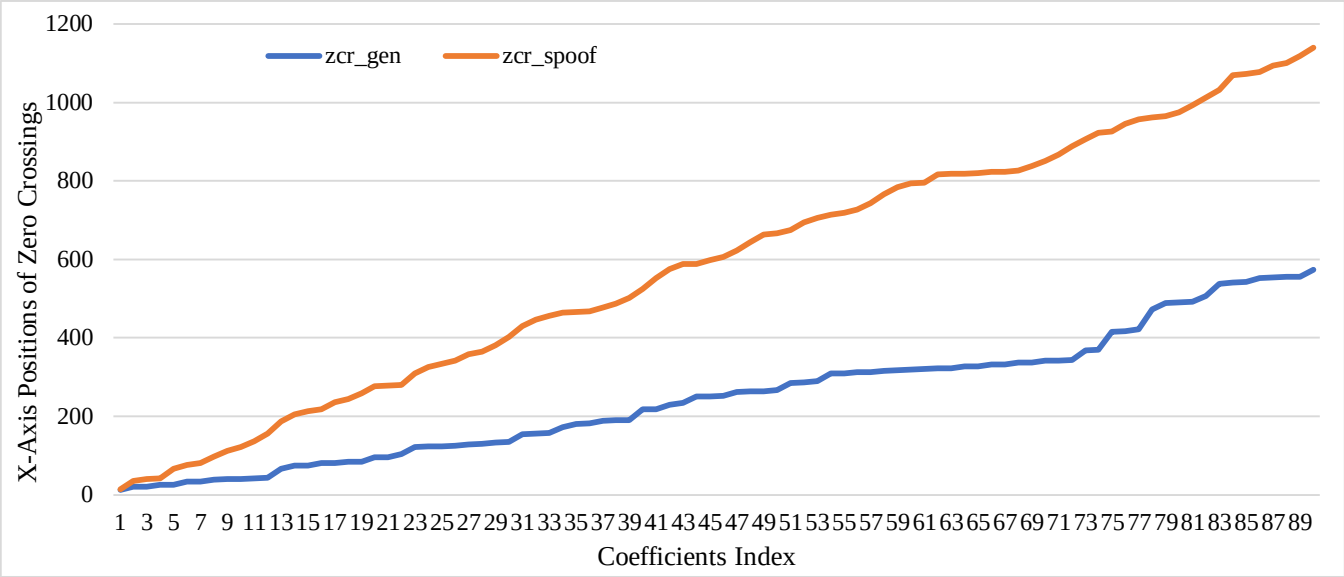

Fig. 14 Zero crossings for genuine and spoof sample

3.4. Timbrel Features

For temporal or time domain features, zero crossings of the signal are analyzed. Zero crossings represent the x-axis positions of zero crossings [45]. This work uses the "zerocros" function provided in the voicebox toolbox. Out of positive and negative crossing sample values, the first ninety coefficients integrate the features mentioned in this work. The feature vector formed is 90*1 dimension. For visualization of zero-crossings, coefficients are compared in the form of a line chart. Figure 14 shows the zero crossings of the genuine and spoofed speech signal. The x-axis positions of zero crossings of spoof speech samples are significantly higher than that of genuine speech. This gives significant non-overlap data representation in case of zero crossings. The feature vector formed using zero-crossings can more efficiently detect reply attacks. The Musical Information Retrieval (MIR) toolbox [46] has timbre audio descriptors that are simple to integrate with the Matlab framework. In [32], Timbrel features roughness, rolloff, irregularity, and brightness demonstrated effective for speaker recognition. In this work, along with the Timbrel features mentioned in [32], several Timbrel features are also evaluated. The Table 6 shows the timbre features used to create a feature vector. This work uses twelve different timbre features to create a feature vector of size twelve.

Table 6. Timbre features used from the MIR toolbox and their description

Sr. No.	Timbre Feature	Description
1	Brightness	It is the midpoint of the frequency energy distribution [32].
2	Entropy	This represents Shanon entropy of the input.
3	Event density	It represents the number of events detected per second.
4	Flatness	It is the ratio of the geometric mean to the arithmetic mean. It represents distribution as smooth or spiky.
5	Inharmonicity	It represents the number of partials, not multiples of the fundamental frequency. It is a value between 0 and 1.
6	Kurtosis	It returns the (excess) kurtosis of the data.
7	Pitch	Mirpitch extract pitches.
8	Irregularity	This represents variations among successive peaks [32].
9	Rolloff	Rolloff is the frequency below which the significant energy (85% or 95%) is concentrated [32]
10	RMS	It represents the global energy of the signal.
11	Skewness	It represents the coefficient of skewness of the data.
12	Spread	It represents the standard deviation of the data

Table 7. Performance of CQCC, MFCC, Integration of CQCC and MFCC, LPC feature extraction techniques [44]

Sr. No.	Method	Number of Coefficients (Rows)	DevEER (%)	EvalEER (%)
1	CQCC	90	11.55	22.8
2	MFCC	30	18.28	34.42
3	LPC without Preprocessing	12	23.56	35.34
4	MFCC + Delta + Double Delta	90	17.08	23.16
5	CQCC + MFCC + Delta + Double Delta	90	10.18	20.96
6	LPC with Preprocessing	12	8.92	35.32

3.5. Integration of Various Features

Implementing most feature extraction techniques is based on a matrix structure consisting of a two-dimensional rectangular array of data elements arranged in rows and columns. The number of rows is nothing but the number of filters or coefficients selected in the feature extraction technique's parameters.

The compatible sizes of matrices are required for concatenation. Horizontal concatenation requires the same number of rows among matrices, and vertical concatenation requires the same number of rows. Hence, this work selects the number of filter coefficients per the matrices' compatibility sizes. If matrix A is $[A]_{m*n}$ and matrix B is $[B]_{m*l}$, then the resultant concatenated matrix will be $[A \text{ concat } B]_{m*(n+l)}$.

Where, m is the number of rows and, n is the number of columns of matrix A, and matrix B is of size m *l, for matrix B, m is the number of rows, l is the number of columns, m is the number of rows, and (n + l) is the number of columns of the resultant concatenated matrix. For example, suppose matrix A = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and matrix B = $\begin{bmatrix} x \\ y \end{bmatrix}$, then concatenation of A & B is [A B] = $\begin{bmatrix} a & b & x \\ c & d & y \end{bmatrix}$.

Based on this mathematical model, different features extracted in this work are concatenated to form a resultant feature vector matrix. For example, as mentioned in, the CQCC feature set with 90-dimension is concatenated with MFCC features of 90-dimension. Here, ninety is the number of rows, representing the number of filter coefficients. Subsequent experimentation in this work follows this approach for concatenating several features.

4. Results

4.1. Results on the Evaluation Set for CQCC, MFCC, and LPC Methods

The work presented in [44] evaluated the performance of CQCC, MFCC, and Integration of CQCC and MFCC of the ASVspoof 2017 version 2 database development set. This work considers the methods presented in [44] on the evaluation set. Table 7 shows the performance of these evaluations. The best cases are shown in bold cases. DevEER (%) presents %EER on the development set in all experiments, and EvalEER (%) offers %EER on the evaluation assigned. Table 7 shows that LPC features are superior with pre-processing steps framing, overlapping, and windowing. Also, LPC features perform better than integrating CQCC and MFCC features on the development set. However, Integrated CQCC and MFCC features perform better on the evaluation set than LPC features.

4.2. LPC and Their Integrations with Cepstral Features

The following experiments integrate LPC features with energy in the residual, CQCC, and MFCC. As mentioned in [40], the LP order is selected as twelve for the conduction of these experiments. Hence, in integrating LPC with CQCC and MFCC, the number of cepstral coefficients determined was 12 for matrix concatenation. Also, in integrating LPC and energy in residual with CQCC and MFCC, the cepstral coefficients selected were 13 for matrix concatenation. Figure 15 shows the performance of these evaluations.

Figure 15 shows that the performance of LPC features slightly improved with the integration of CQCC and MFCC features on the development set with an EER of 8.25%. On the evaluation set, integration of LPC features, energy in residual, MFCC, and CQCC resulted in better performance with EER of 33.97% than other combinations. Then, LPCC features are evaluated on the database development and evaluation set.

4.3. LPCC and Their Integrations with Cepstral Features

Next, LPCC features are integrated with energy in the residual, CQCC, and MFCC. Similar to experiments carried out for LPC and their integrations, LPCC features were considered 12. Therefore, in integrating LPCC with CQCC and MFCC, the number of cepstral coefficients selected was 12 for matrix concatenation. Also, in integrating LPCC and energy in residual with CQCC and MFCC, the cepstral coefficients determined were 13 for matrix concatenation. The results of these experiments are shown in Figure 16.

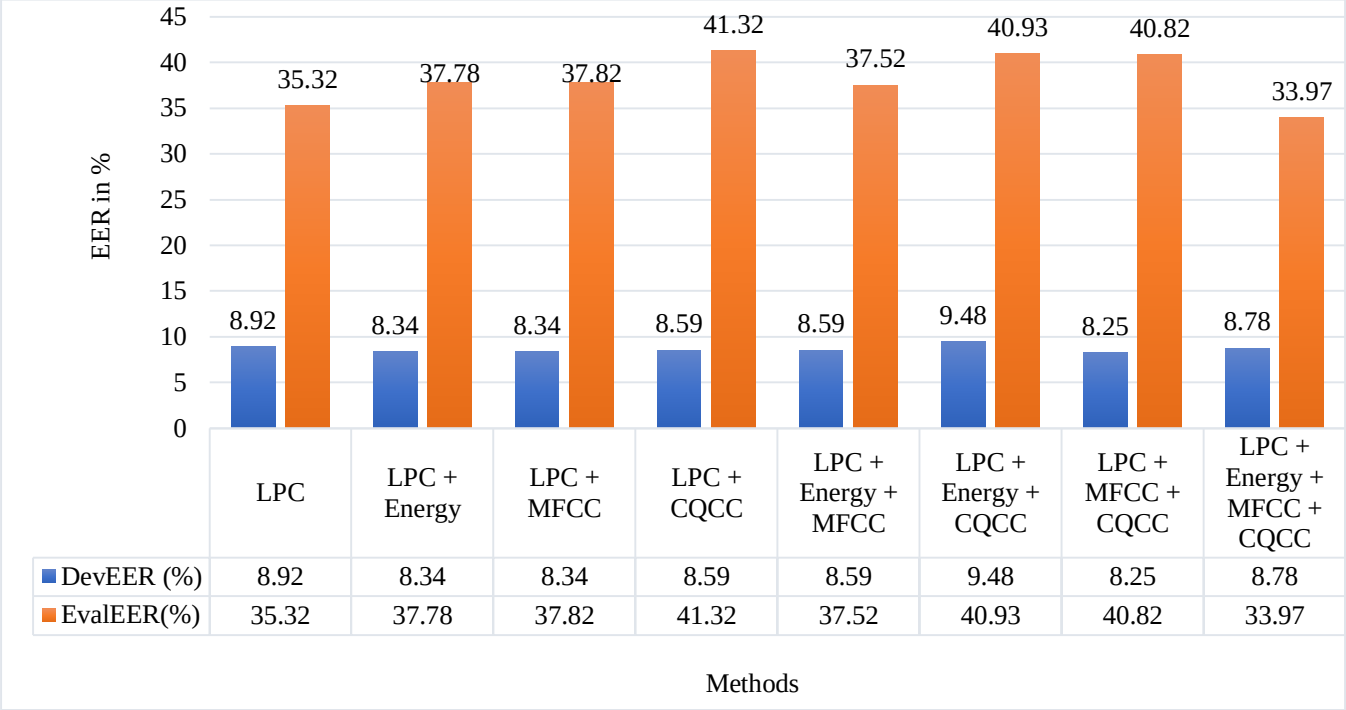

Fig. 15 Performance of integration of LPC with energy in the residual, CQCC and MFCC feature extraction techniques

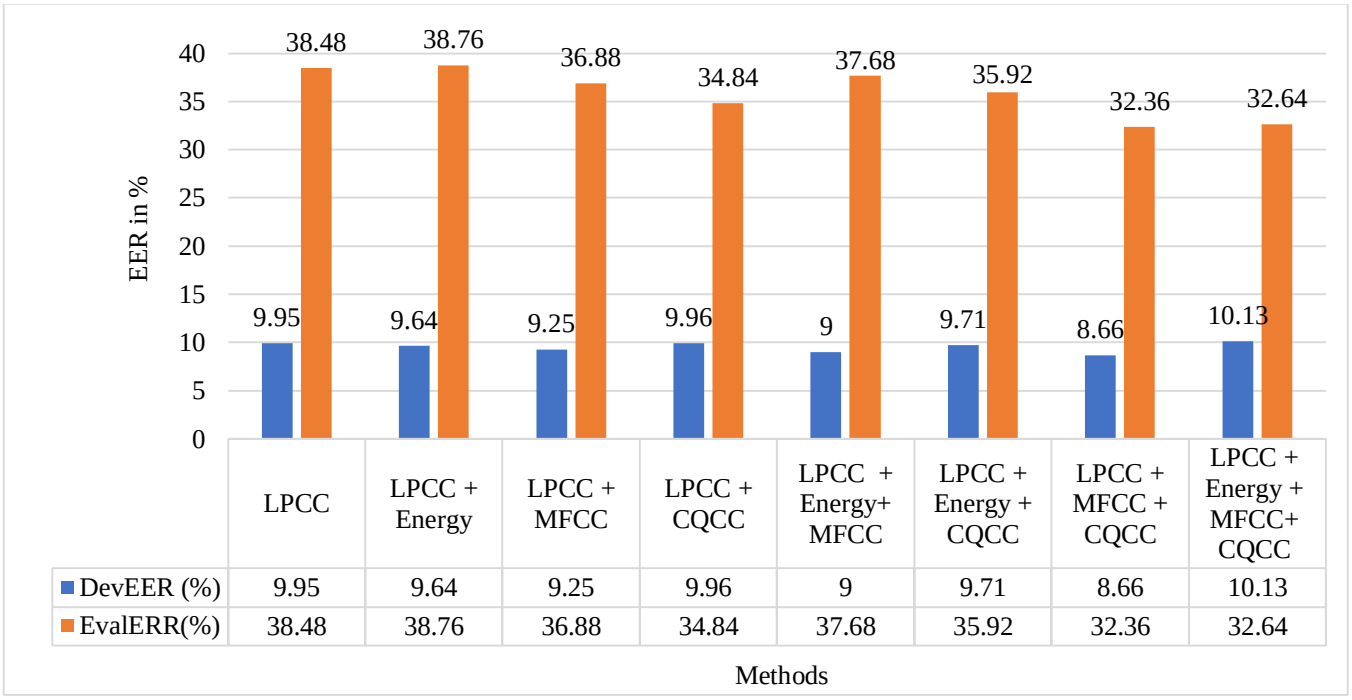

Fig. 16 Performance of integration of LPCC with energy in the residual, CQCC and MFCC feature extraction techniques

Figure 16 shows that the performance of LPCC features, when integrated with CQCC and MFCC, improves on the evaluation set compared to that of LPC features. However, LPCC features performance is slightly degraded on the development set. It can be inferred that LPCC features are more suitable for the evaluation set, and LPC features are ideal for the development set. However, the performance integrations of CQCC and MFCC features (EER 20.96% mentioned in Table 7) are better than integrating LPCC features and their integrations with other cepstral features and energy in residual. Also, it has been observed that energy in the residual extracted from the "lpcauto" function does not

show much improvement because of which, in the next experiment, it is not integrated with any feature extraction schemes such as LFCC, Timbrel feature and ZCR.

4.4. LFCC and Their Integrations of Other Cepstral Features and Linear Prediction-Based Features

Further analysis is carried out with the LFCC feature extraction technique. Baseline LFCC features provided by ASVspoof 2019 organizers are evaluated on the ASVspoof 2017 version 2 development and evaluation set. To observe the performance, LFCC features are also integrated with CQCC, MFCC, LPC, and LPCC features.

From the results obtained till Figure 16, it is evident that cepstral features perform better with 90 coefficients. Hence, the number of cepstral coefficients for CQCC, MFCC, and LFCC was considered 90. Also, the integration of ninety cepstral features with ninety linear predication-based coefficients was carried out. Figure 17 shows the performance of LFCC features and their integration with other features.

It is observed that baseline LFCC features perform better with EER of 7.44% than CQCC features, MFCC features, LPC features and several previously mentioned integrations on the development set. Further experiments are carried out by integrating LFCC with CQCC, MFCC, LPC, and LPCC features. As shown in Figure 17, the integration of LFCC with CQCC, MFCC, and LPC resulted in an EER of 7.2% on the development set, and the integration of LFCC, CQCC, MFCC, and LPCC resulted in an EER of 20.51% on the evaluation set.

From the results observed till now, it is evident that integrating multiple features improves the performance.

4.5. Integration of Timbrel Feature Set with Cepstral Features and Linear Prediction-Based Features

The performance of the feature vector formed using the timbre features mentioned in Table 6 is evaluated by integrating with all feature extraction schemes used in this work. Twelve features mentioned in Table 7 are concatenated, and a feature vector is formed. This Timbrel feature vector is concatenated with various combinations of CQCC, MFCC, LFCC, LPC and LPCC features. The number of cepstral and linear prediction-based features for matrix concatenation was considered 12. Table 8 shows the performance of the development and evaluation set. From Table 8, it is observed that the Timbrel feature set formed using several features mentioned in Table 6, when integrated with cepstral features and linear prediction-based features, has scope for further improvement.

4.6. Integration of Zero-Cross Rate Timbrel Feature with Cepstral Features and Linear Prediction-Based Features

Next, experiments are carried out by integrating Zero Crossings (ZCR), which represent the number of times the signal crosses the x-axis and is analyzed with cepstral domain features and linear prediction-based features. Zero crossings are integrated with CQCC, MFCC, LPC, LPCC and LFCC features. Ninety zero-cross rate coefficients were considered for integrations with ninety cepstral and linear prediction-based coefficients.

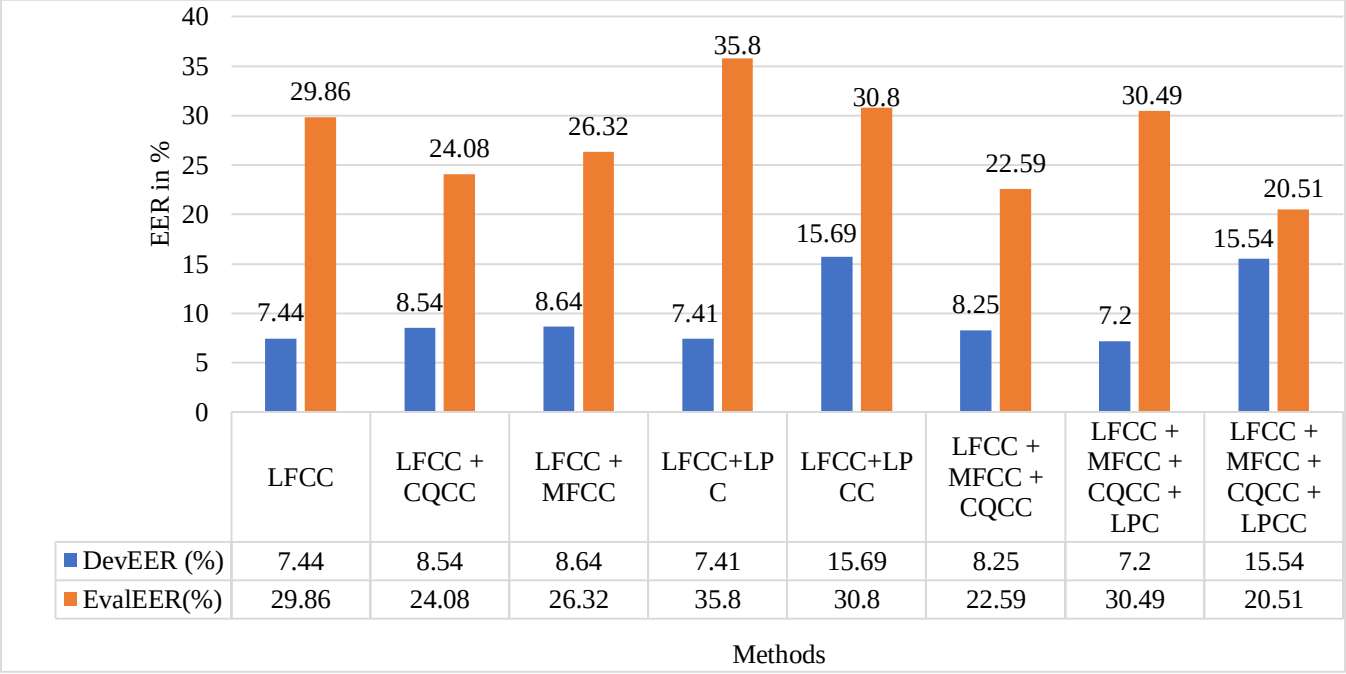

Fig. 17 Performance of LFCC features and integration of LFCC features with CQCC, MFCC, LPC, and LPCC features

Table 8. Performance evaluation of integration of zero crossings with CQCC, MFCC, LFCC, LPC, and LPCC

Sr. No.	Method	Number of Coefficients (Rows)	DevEER (%)	EvalEER (%)
1	CQCC + Timbre	12	9.72	34.94
2	MFCC + Timbre	12	18.81	36.63
3	CQCC + MFCC +Timbre	12	9.17	33.58
4	CQCC + MFCC + LPC + Timbre	12	8.8	34.96
5	LPC + Timbre	12	8.43	35.84
6	LPC + CQCC + Timbre	12	7.52	35.31
7	LPC + MFCC + Timbre	12	8.14	36.65
8	LPCC + Timbre	12	10.11	38.59
9	LPCC + CQCC + Timbre	12	8.75	37.38
10	LPCC + MFCC + Timbre	12	10.49	37.38
11	CQCC + MFCC + LPCC + Timbre	12	8.58	34.51
12	LFCC + Timbre	12	8.68	37.56
13	LFCC + CQCC + Timbre	12	9.48	35.52
14	LFCC + MFCC + Timbre	12	10.76	34.76
15	LFCC + CQCC + MFCC + Timbre	12	9.77	33.88
16	LFCC + LPC + Timbre	12	7.45	36.23
17	LFCC + CQCC + LPC + Timbre	12	7.38	35.2
18	LFCC + MFCC + LPC + Timbre	12	8.11	35.68
19	LFCC + CQCC + MFCC + LPC + Timbre	12	7.56	34.16
20	LFCC + LPCC + Timbre	12	8.63	37.81
21	LFCC + CQCC + LPCC + Timbre	12	8.05	36.14
22	LFCC + MFCC + LPCC + Timbre	12	9.97	36.97
23	LFCC + CQCC + MFCC + LPCC + Timbre	12	10.01	31.27

Table 9. Comparison of achieved results with recent approaches

Sr. No.	Authors	Method	DevEER (%)	EvalEER (%)
1	Gupta et al., 2023 [4]	Score level fusion of CQCC, CFCC, CFCCIF, CFCCIF-ESA, CFCCIF-QESA	9.21	11.24
2	Kamble and Patil 2021, [10]	CQCC + LFCC + MFCC + TECC	6.68	10.45
3	Current Authors	LFCC + MFCC + CQCC + LPC	7.2	30.49
4	Current Authors	LFCC + MFCC + CQCC + LPCC	15.54	20.51
5	Current Authors	ZCR + LFCC	6.2	28.88
6	Current Authors	ZCR + CQCC + MFCC	9.08	19.97
7	Current Authors	ZCR + LPC	5.44	31.9
8	Current Authors	ZCR + MFCC + CQCC + LFCC + LPCC	14.37	17.79

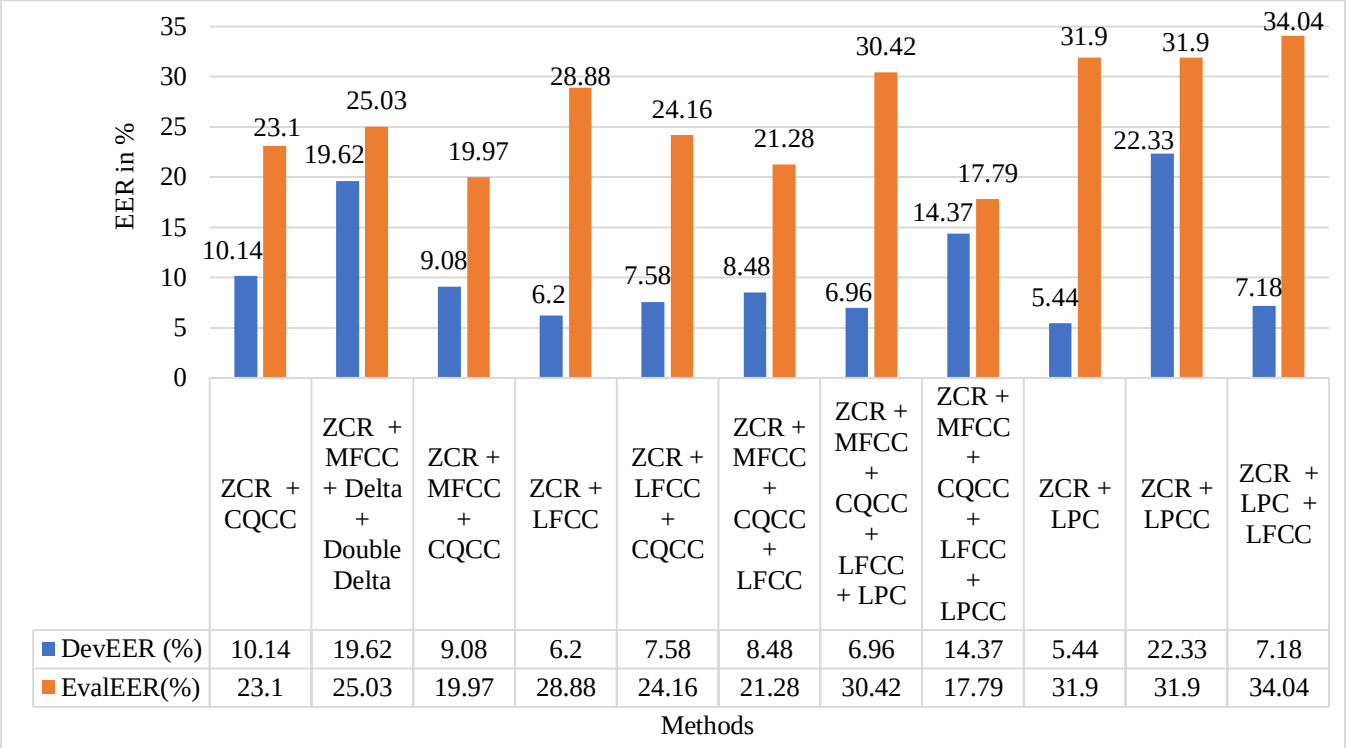

Fig. 18 Performance evaluation of integration of zero crossings with CQCC, MFCC, LFCC, LPC, and LPCC

Figure 18 shows these integrations along with performance on the development and evaluation set. From Figure 18, EER 6.2% is achieved on the development set for Zero Cross Rate (ZCR) and LFCC integrations and 19.97% on the evaluation set for ZCR, MFCC, and CQCC.

It has been observed that zero crossings integrated with LPC coefficients resulted in better performance with an EER of 5.44% on the development set, whereas zero crossings integrated with CQCC, MFCC, LPCC, and LFCC resulted in an EER of 17.79% on the evaluation set. These results show the effectiveness of zero crossings in detecting replay attacks.

5. Discussion

From the literature, it is evident that score-level fusion and feature fusion approaches are widely preferred for replay attack detection. Also, the use of multiple features and combining them to achieve improved performance is state-of-the-art in the area of countermeasures for replay attack detection. This work mainly focused on the performance evaluation of multiple cepstral and linear-prediction-based features.

Many past approaches have been presented with a focus on score-level fusion of multiple features. This work mainly focused on integrating multiple features via serial fusion. This work evaluated the Timbrel features and found the effectiveness of zero-cross-rate timbre features. The results of this work demonstrated the effectiveness of cepstral features such as LFCC, linear predictive coding feature coefficients and Zero-cross rate. It is evident from the results that LPC features are effective in the development set, whereas LPCC features are effective in the evaluation set of ASVspoof 2017 version 2. Table 9 presents the comparison of achieved results for the best cases of integration of cepstral features, linear predication-based features, and Timbrel features zero cross rate with recent approaches from the literature.

From Table 9, it is observed that on the development set, proposed integration approaches highlighted in bold cases performed better than some recent approaches. EER of 5.44% on the development set for zero-cross rate and LPC integration. This indicates that the serial feature fusion approach also achieves better results.

LPCs have more decorrelation, representing more non-overlapping representation (Figure 12) of genuine and spoofed speech signals. Also, zero crossings of genuine and spoofed speech signals show more differentiation (Figure 14).

This differentiation has improved performance for detecting spoofed signals from genuine signals. On the evaluation set, better EER 17.79% is achieved for integration of zero-cross rate, MFCC, CQCC, LFCC, and LPCC, among others mentioned in this work. There is a scope for improvement on the evaluation set, and future work will focus on carrying score level fusion for multiple features such as teager coefficients.

6. Conclusion

This work has evaluated the performance of CQCC, MFCC, LFCC, LPC, and LPCC feature extraction schemes for detecting reply attacks on ASV systems. These evaluations are carried out on the ASVspoof 2017 version 2 database. Among these baseline methods, LFCC features performed better on the development set with an EER of 7.44%, and CQCC features showed better performance on the evaluation set with an EER of 22.8%. Using multiple features and a feature fusion approach for improved performance is the state-of-the-art for detecting spoofing attacks. In this work, multiple features, namely, CQCC, MFCC, LFCC, LPC, and LPCC, are integrated, and the performance of various integrations is evaluated. On the development set, the integration of LFCC, MFCC, CQCC, and LPC features resulted in an EER of 7.41% and on the evaluation set, the integration of LFCC, MFCC, CQCC, and LPCC features resulted in an EER of 20.51%.

Further, Timbrel features zero cross rate are integrated considering various combinations with cepstral features and LPC. An EER of 5.44% is achieved on the development set for the integration of zero cross rate and LPC feature, and an EER of 17.79% is achieved on the evaluation set for the integration of zero cross rate, MFCC, CQCC, LFCC, and LPCC features. It found that integrating zero-crossings performed better than integrating the feature vector formed using various Timbrel features with cepstral features and linear prediction-based features. The effectiveness of Timbrel features, i.e., zero crossings for replay attack detection, is validated from the results achieved. Also, the results show that LFCC and LPC features are more effective in the development set, and CQCC and LPCC features are effective in the evaluation set. This work inferred that multiple different features may show improved performance for every database. Also, the serial feature fusion approach can be suitable for replay attack detection. In future, the effectiveness of integrating cepstral coefficients based on teager energy with Timbrel features can be tested along with score level fusion.

Acknowledgement

Authors are thankful to ASVspoof 2017 and 2019 challenge organizers for providing database and Matlab-based reference replay attack spoofing detection. Also, the authors are thankful to researchers who have provided publicly available toolsets such as the Voicebox toolbox, MIR toolbox, Bosaris toolkit, VLFeat library and PRAAT.

References

[1] Madhu R. Kamble et al., "Advances in Anti-Spoofing: From the Perspective of ASVspoof Challenges," *APSIPA Transactions on Signal and Information Processing*, vol. 9, pp. 1-18, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] A.K. Jain, A. Ross, and S. Pankanti, "Biometrics: A Tool for Information Security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125-143, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[3] Anamika Baradiya, and Vinay Jain, "Speech and Speaker Recognition Technology Using MFCC and SVM," *SSRG International Journal of Electronics and Communication Engineering*, vol. 2, no. 5, pp. 6-9, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[4] Priyanka Gupta, Piyushkumar K. Chodingala, and Hemant A. Patil, "Replay Spoof Detection Using Energy Separation Based Instantaneous Frequency Estimation from Quadrature and In-Phase Components," *Computer Speech & Language*, vol. 77, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nicholas Evans, Tomi Kinnunen, and Junichi Yamagishi, "Spoofing and Countermeasures for Automatic Speaker Verification," *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 925-929, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[6] Zhizheng Wu et al., "ASVspoof 2015: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan," *Training*, pp. 1-5, 2014. [Google Scholar]

[7] Tomi Kinnunen et al., "ASVspoof 2017: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan*," *Taining*, pp. 1-6, 2018. [Google Scholar]

[8] Andreas Nautsch et al., "ASVspoof 2019: Spoofing Countermeasures for the Detection of Synthesized, Converted and Replayed Speech," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 252-265, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Héctor Delgado et al., "ASVspoof 2021: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan," *Electrical Engineering and Systems Science*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Madhu R. Kamble, and Hemant A. Patil, "Detection of Replay Spoof Speech Using Teager Energy Feature Cues," *Computer Speech & Language*, vol. 65, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Roberto Font, Juan M. Espín, and María José Cano, "Experimental Analysis of Features for Replay Attack Detection-Results on the ASVspoof 2017 Challenge," *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 7-11, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[12] Hemant A. Patil et al., "Novel Variable Length Teager Energy Separation Based Instantaneous Frequency Features for Replay Detection," *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 12-16, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Sarfaraz Jelil et al., "Spoof Detection Using Source, Instantaneous Frequency and Cepstral Features," *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 22-26, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] Marcin Witkowski et al., "Audio Replay Attack Detection Using High-Frequency Features," *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 27-31, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Choon Beng Tan et al., "A Survey on Presentation Attack Detection for Automatic Speaker Verification Systems: State-of-the-Art, Taxonomy, Issues and Future Direction," *Multimedia Tools and Applications*, vol. 80, pp. 32725-32762, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Xiaojiang Pen et al., "Bag of Visual Words and Fusion Methods for Action Recognition: Comprehensive Study and Good Practice," *Computer Vision and Image Understanding*, vol. 150, pp. 109-125, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Kamer Vishi, and Vasileios Mavroeidis, "An Evaluation of Score Level Fusion Approaches for Fingerprint and Finger-Vein Biometrics," *Computer Science*, pp. 1-10, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Quan-Sen Sun et al., "A New Method of Feature Fusion and Its Application in Image Recognition," *Pattern Recognition*, vol. 38, no. 12, pp. 2437-2448, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[19] Cemal Hanilci, "Speaker Verification Anti-Spoofing Using Linear Prediction Residual Phase Features," *2017 25th European Signal Processing Conference (EUSIPCO)*, Greece, pp. 96-100, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[20] Madhu R. Kamble, and Hemant A. Patil, "Novel Energy Separation Based Frequency Modulation Features for Spoofed Speech Classification," *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jichen Yang, Rohan Kumar Das, and Haizhou Li, "Extended Constant-Q Cepstral Coefficients for Detection of Spoofing Attacks," *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, USA, pp. 1024-1029, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[22] Krishna Dutta, Madhusudan Singh, and Debadatta Pati, "Detection of Replay Signals Using Excitation Source and Shifted CQCC Features," *International Journal of Speech Technology*, vol. 24, pp. 497-507, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Leian Liu, and Jichen Yang, "Study on Feature Complementarity of Statistics, Energy, and Principal Information for Spoofing Detection," *IEEE Access*, vol. 8, pp. 141170-141181, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Madhu R. Kamble, Hemlata Tak, and Hemant A. Patil, "Amplitude and Frequency Modulation-Based Features for Detection of Replay Spoof Speech," *Speech Communication*, vol. 125, pp. 114-127, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] B.T. Balamurali et al., "Toward Robust Audio Spoofing Detection: A Detailed Comparison of Traditional and Learned Features," *IEEE Access*, vol. 7, pp. 84229-84241, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[26] Khomdet Phapatanaburi et al., "Replay Attack Detection Using Linear Prediction Analysis-Based Relative Phase Features," *IEEE Access*, vol. 7, pp. 183614-183625, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Madhusudan Singh, and Debadatta Pati, "Usefulness of Linear Prediction Residual for Replay Attack Detection," *AEU - International Journal of Electronics and Communications*, vol. 110, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[28] Madhusudan Singh, and Debadatta Pati, "Combining Evidences from Hilbert Envelope and Residual Phase for Detecting Replay Attacks," *International Journal of Speech Technology*, vol. 22, pp. 313-326, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Zeyan Oo et al., "Replay Attack Detection with Auditory Filter-Based Relative Phase Features," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[30] Sang Kyoon Park et al., "Zero-Crossing-Based Feature Extraction for Voice Command Systems Using Neck-Microphones," *International Symposium on Neural Networks in Advances in Neural Networks - ISNN 2007*, pp. 1318-1326, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[31] V.M. Sardar, and S.D. Shirbahadurkar, "Speaker Identification of Whispering Speech: An Investigation on Selected Timbrel Features and KNN Distance Measures," *International Journal of Speech Technology*, vol. 21, pp. 545-553, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[32] Vijay M. Sardar, and S.D. Shirbahadurkar, "Timbre Features for Speaker Identification of Whispering Speech: Selection of Optimal Audio Descriptors," *International Journal of Computers and Applications*, vol. 43, no. 10, pp. 1047-1053, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] Vijay M. Sardar, Manisha L. Jadhav, and Saurabh H. Deshmukh, "Use of Median Timbre Features for Speaker Identification of Whispering Sound," *Techno-Societal 2020*, pp. 31-41, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[34] Héctor Delgado et al., "ASVspoof 2017 Version 2.0: Meta-Data Analysis and Baseline Enhancements," *Odyssey 2018 - The Speaker and Language Recognition Workshop*, Les Sables d'Olonne, France, 2018. [Google Scholar] [Publisher Link]

[35] Tomi Kinnunen et al., "The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection," *Proceedings of the Annual Conference of the International Speech Communication Association*, pp. 2-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[36] Massimiliano Todisco, Héctor Delgado, and Nicholas Evans, "Constant Q Cepstral Coefficients: A Spoofing Counter Measure for Automatic Speaker Verification," *Computer Speech & Language*, vol. 45, pp. 516-535, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[37] Prashanth Kannadaguli, and Vidya Bhat, "Phoneme Modeling for Speech Recognition in Kannada Using Multivariate Bayesian Classifier," *SSRG International Journal of Electronics and Communication Engineering*, vol. 1, no. 9, pp. 1-4, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[38] Sabur Ajibola Alim, and Nahrul Khair Alang Rashid, *Some Commonly Used Speech Feature Extraction Algorithms*, Natural to Artificial Intelligence - Algorithms and Applications, IntechOpen, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[39] K.M. Ravikumar, H.C. Nagaraj, and R. Rajagopal, "An Approach for Objective Assessment of Stuttered Speech Using MFCC Features," *The International Congress for Global Science and Technology*, vol. 19, 2009. [Google Scholar] [Publisher Link]

[40] Linqiang Wei et al., "New Acoustic Features for Synthetic and Replay Spoofing Attack Detection," *Symmetry*, vol. 14, no. 2, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[41] Tae Hong Park, *Towards Automatic Musical Instrument Timbre Recognition [Microform]*, ProQuest Dissertations and Thesis, Princeton University, 2004.

[42] Anthony Larcher, Sylvain Meignier, and Kong Aik Lee, SIDEKIT Documentation, 2020. [Online]. Available: https://projets-lium.univ-lemans.fr/sidekit/_downloads/544a50fdcc0129b614b6f4b90f1c89d0/sidekit.pdf

[43] Paul Boersma, and David Weenink, Praat: Doing Phonetics by Computer. [Online]. Available: http://www.praat.org

[44] K. Raja, "Detection and Prevention of Ransomware Attacks using AES and RSA Algorithms," *DS Jornal of Digital Science and Technology*, vol. 1, no. 1, pp. 1-9, 2022. [CrossRef] [Publisher Link]

[45] Mike Brookes, VOICEBOX: Speech Processing Tool Box for MATLAB. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[46] Olivier Lartillot, Petri Toiviainen, and Tuomas Eerola, "A Matlab Toolbox for Music Information Retrieval," *Data Analysis, Machine Learning and Applications Conference*, pp. 261-268, 2008. [CrossRef] [Google Scholar] [Publisher Link]