IJIST
P ISSN : 2618-1630
E ISSN : 2909-6130

Original Article

Voice Spoofing Countermeasure Based on Spectral Features to Detect Synthetic Attacks Through LSTM

Gulam Qadir[1], Saima Zareen[1], Farman Hassan[1], Auliya Ur Rahman[1]
[1]University of Engineering and Technology Taxila, Punjab Pakistan.
*Correspondence: Farman Hassan, Email ID: farmanhassan555@gmail.com.

Citation | Gulam Qadir, Saima Zareen, Farman Hassan, and Auliya Ur Rahman. 2022. "Voice Spoofing Countermeasure Based on Spectral Features to Detect Synthetic Attacks Through LSTM". International Journal of Innovations in Science & Technology 3 (4):153-165. https://journal.50sea.com/index.php/IJIST/article/view/124.
DOI | https://doi.org/10.33411/IJIST/2021030512
Received | Dec 15, 2021; **Revised** | Dec 29, 2021 **Accepted** | Jan 06, 2022; **Published** | Jan 07, 2022.

With the growing number of voice-controlled devices, it is necessary to address the potential vulnerabilities of automatic speaker verification (ASV) against voice spoofing attacks such as physical access (PA) and logical access (LA) attacks. To improve the reliability of ASV systems, researchers have developed various voice spoofing countermeasures. However, it is hard for the voice anti-spoofing systems to effectively detect the synthetic speech attacks that are generated through powerful spoofing algorithms and have quite different statistical distributions. Most importantly, the speedy improvement of voice spoofing structures is producing the most effective attacks that make ASV structures greater vulnerable to stumble on those voice spoofing assaults. in this paper, we proposed a unique voice spoofing countermeasure this is successful to hit upon the LA attacks (i.e., artificial speech and transformed speech) and classify the spoofing structures by the usage of long short-term reminiscence (LSTM). The novel set of spectral features i.e., Mel-frequency cepstral coefficients (MFCC), gammatone cepstral coefficients (GTCC), and spectral centroid are capable to seize maximum alterations present in the cloned audio. The proposed system achieved remarkable accuracy of 98.93%, precision of 100%, recall of 92.32%, F1-score of 96.01%, and an equal error rate (EER) of 1.30%. Our method achieved 8.5% and 7.02% smaller EER than the baseline methods such as Constant-Q cepstral coefficients (CQCC) using Gaussian mixture model (GMM) and Linear frequency cepstral coefficients (LFCC) using GMM, respectively. We evaluated the performance of the proposed system on the standard dataset i.e., ASVspoof2019 LA. Experimental results and comparative analysis with other existing state-of-the-art methods illustrate that our method is reliable and effective to be used for the detection of voice spoofing attacks.

Keywords: ASVspoof2019 LA dataset; Deep Learning; Spoofing countermeasure; Synthetic Speech; Voice anti-spoofing.

Acknowledgment.
We are grateful to the ASV spoof organizers for providing the dataset.
Project details. NIL

Author's Contribution.
The corresponding author equally contributed to this work.

Conflict of interest.
Authors claim that there exists no conflict of interest for publishing this manuscript in IJIST.

Introduction

ASV verifies the identity of users based on the voice presented to the ASV systems. Within the previous couple of years, we've witnessed a rapid evolution in voice biometrics-primarily based on user authentication. ASV structures are embedded in several gadgets inclusive of smart speakers (Amazon Alexa, Google home), and smartphones for the authentication in diverse software domains i.e., e-trade, banking structures, home automation, and special utility logins [1]. Google domestic gets speech commands from the users and performs several features which include putting reminders, text or name, remaining, and starting doorways, and unlocking cell telephones [2]. These applications are based on ASV systems [3]. Banking systems are also using voice-based authentication systems to verify customers such as BBVA's and Barclays Wealth have been using voice biometrics for verification of telephone callers. The Grant bank has developed a voice-based application, which allows customers to perform transactions simply by using voice commands [4].

We have witnessed an exponential and rapid growth in voice-driven authentication systems due to the Covid-19 pandemic. Social distancing and lockdown have limited the capability of facial, fingerprint, or iris recognition. This pandemic has urged the world to shift the verification measures based on a human to machine and human to human interactions to voice-based authentication systems [32]. Consequently, voice-primarily based authentication has emerged as a most possible and simple answer than every other biometrics approach which includes the iris, facial, and fingerprint. Moreover, voice-based authentication is considered to be the most economical and efficient than other biometrics methods [33]. ASV is playing a significant role in the biometric verification process. ASV uses the acoustic features of a person to authenticate the users [5]. However, intruders can mislead ASV systems by voice spoofing attacks i.e., voice replay, mimics or twins, voice conversion (VC), and syntenic speech [6,7]. Among these spoofing attacks, synthetic spoofing attacks (text-to-speech (TTS) and VC are threats to ASV systems that occur due to the rapid development of synthetic methods [8,9].

Stand-alone voice anti-spoofing techniques are developed to enhance the security and reliability of ASV systems. ASVspoof challenge series has been providing datasets [7,10,11] and provided their standard metrics for voice anti-spoofing speaker verification. In this work, we cognizance of voice anti-spoofing la attacks which include detecting true and spoofed speech produced with the aid of VC and TTS spoofing structures, detecting unseen attacks, and type of acknowledged assaults. Traditional techniques focus on feature engineering and hand-crafted features i.e., Cochlear filter cepstral coefficients (CFCCIF) [12], Linear cepstral coefficients (LFCC) [13], and Constant-Q cepstral coefficients (CQCC) [14] have shown better results against spoofing attacks. GMM is used in traditional methods [12-16] as a backend classifier.

The research community has also explored various deep learning methods [17-24] to detect LA attacks. In [17], deep learning models were investigated for anti-spoofing. Combining Convolutional Neural Network (CNN) and Recurrent Neural Network (RNNs) showed robustness against spoofing attacks. In [18], a deep residual community (ResNet) turned into used with temporal pooling. In [19], ResNet becomes employed with the most margin of cosine loss and frequency masks augmentation. In [20], mild convolutional gate RNN become adopted to enhance the lengthy-time period dependency for the detection of voice spoofing assaults. In [21], a technique of characteristic genuinization based on a mild CNN machine changed into a proposed that performed properly towards l. a. attacks. In [21], a technique of feature genuinization based on a light CNN system was proposed that performed well against LA attacks. In [22], the transfer learning approach was explored with the ResNet network. The research community introduced model fusion based on sub-band modeling [23] and various features [24,26] to enhance the performance of voice anti-spoofing systems. In[27], three features such as MFCC, CQCC, and short-term Fourier transform (STFT) were integrated to detect voice spoofing attacks detection. ResNet was employed for classification purposes. It was observed that these three different variants of ResNet with MFCC, CQCC, and STFT produce better results than the baselines (CQCC-GMM [7] and LFCC-GMM [7]). In [28], two features such as CQCC and log power magnitude were used to design spoofing countermeasure. A deep neural network (DNN) was employed to discriminate the spoof and authentic audio. The DNN was based on a squeeze-excitation network and residual networks. This [28] framework yielded better results than the existing state-of-the-art methods, but the fusion of squeeze-excitation network and residual network substantially increased the training time. Few works [29,30,31] have used machine-learned features. In [29], two different deep learning models such as light convolutional neural network (LCNN) and gated recurrent neural network (GRNN) were used as features extractors from bonafide and spoof samples. Extracted functions had been used to train three gadget mastering algorithms which include SVM, linear discriminant analysis, and its probabilistic version (PLDA) for discriminating spoof and legitimate audio. In [30], DNN becomes hired to generate body-level posteriors and bottleneck capabilities to distinguish spoof and authentic audio. In [31], RNN based sequence level and DNN based frame-level features were used to design spoofing countermeasures for LA attacks. Different machine learning algorithms such as LDA, SVM, and gaussian density function were employed for classification

purposes. Three model structures such as stacked autoencoder, multi-task joint learned DNN, and spoofing discriminant DNN was used for DNN based frame-level features. LSTM-RNN and BiLSTM-RNN were used for RNN based sequence-level features. These techniques achieved better classification results, but the computation cost was maximum.

In this work, we proposed a novel set of integrated spectral features for a voice anti-spoofing framework that improves the detection of unseen attacks, synthetic speech and classifies the cloning algorithms. The proposed system is robust to capture the alterations produced by the spoofing systems in speech signals. Experimental results show that our method outperformed all other existing state-of-the-art methods on the ASV spoof 2019 LA dataset. The main contributions of our work are as under:

We proposed a novel set of integrated features, which better capture the maximum distortions created by the voice spoofing algorithms and traits of the speaker-induced variations in the genuine audio. The proposed system is capable of successfully classifying the spoofing systems and detecting unseen LA attacks.Our method is capable to improve the security of automatic speaker verification systems against speech synthetic and voice converted attacks. We performed rigorous experiments on the ASVspoof2019 LA dataset to show the significance of the proposed system for the detection of LA attacks.

The remaining paper is organized as, in section 2, we discuss the proposed methodology followed by the feature extraction and classification. Section 3 presents the detail of experimental results and discussion. Finally, we conclude our work in section 4.

PROPOSED METHODOLOGY

This section presents a detailed description of the proposed voice spoofing counter measure. The main objective of the proposed framework is to discriminate authentic and cloned audio, classify spoofing systems, and detect unseen LA attacks. The proposed system comprises of two stages such as features extraction and classification. In the initial stage, the proposed spoofing countermeasure system takes audio as input and extracts three features i.e., 14-dim MFCC, 14-dim GTCC, and 1-dim spectral centroid. In the second stage, we employed LSTM for classification purposes. The detailed working mechanism of the proposed system is shown in Figure 1.

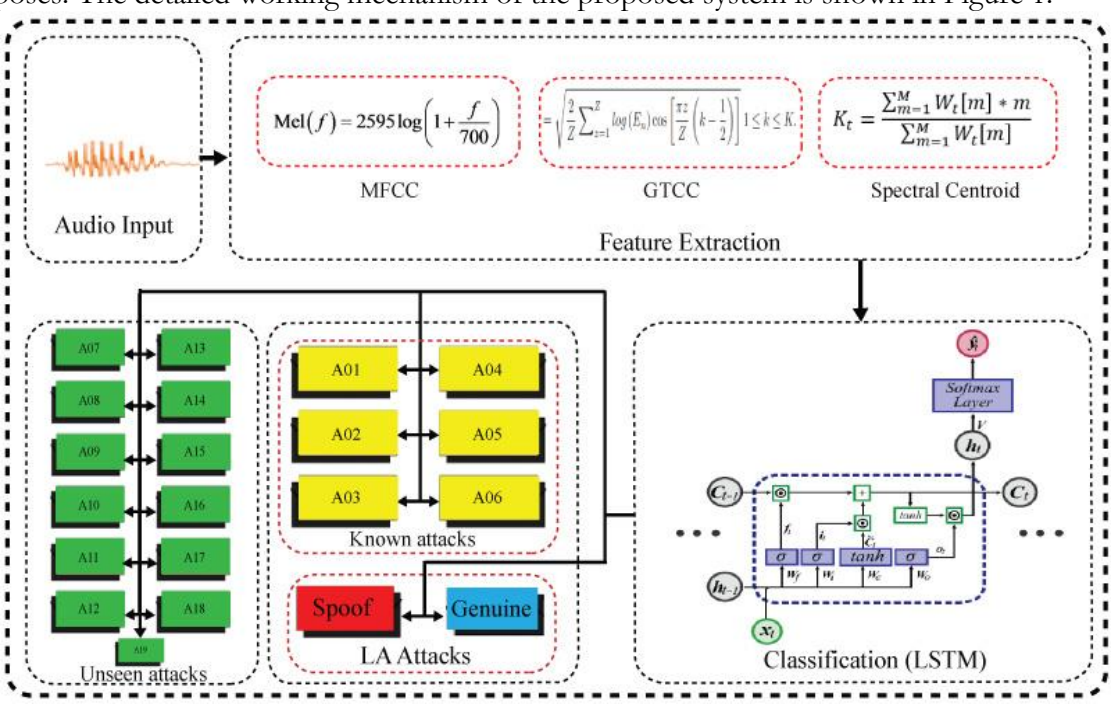

Figure 1. Proposed System.

A. Feature Extraction

To develop a robust voice spoofing countermeasure, we need to propose a robust audio feature descriptor that is capable of capturing the algorithmic artifacts from the spoof speech signals. For this purpose, we proposed a novel set of integrated spectral features that can extract highly discriminative information from the audio signals to accurately detect LA attacks, classify the spoofing systems, and detect the unseen voice spoofing attacks. The detailed feature extraction process is discussed below:

a) MFCC

MFCC takes voice as an input, and it calculates cepstral coefficients from it. The stepwise computation of MFCC from audio is explained in the below Figure 2. We extracted 14-dim MFCC features from audio by pre-emphasis. Pre-emphasis is used to compensate for the high-frequency part of the speech signal followed by the frame blocking in which speech signals are segmented into frames of 15-20ms to investigate the speech over a brief period. Next, each frame of audio is multiplied by the hamming window to keep the continuity between the frames, enhance harmonics, and minimize the edge effects. In the next step, we employed the Fast Fourier Transform (FTF) to

get the magnitude spectrum of each frame. The power spectrum obtained after employing FFT is then mapped to mel-scale. Next, we multiplied the magnitude frequency by 40 triangular bandpass filters to get the log energy.

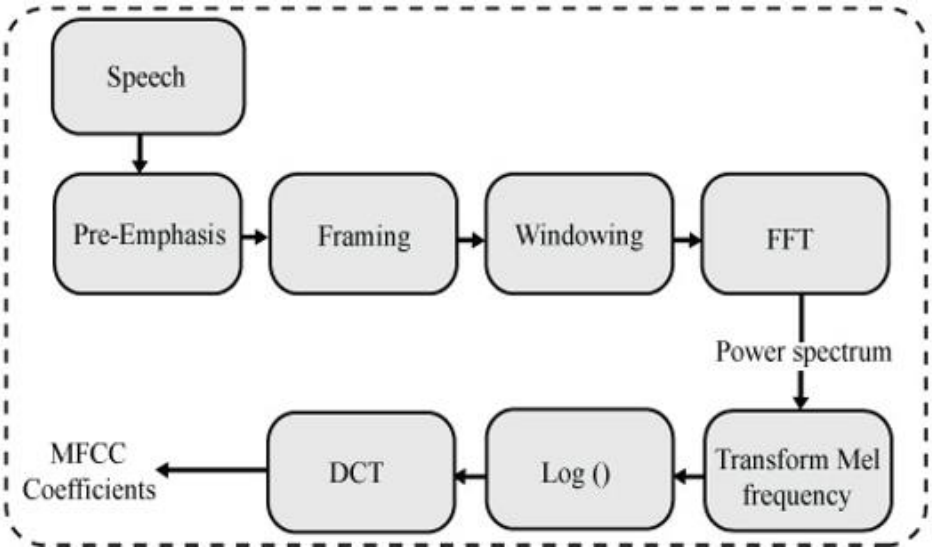

Figure 2. MFCC feature extraction process.

Later, we computed Discrete Fourier Transform (DFT) of the list of Mel log powers. Finally, the amplitude of the spectrums is selected as MFCCs.

b) GTCC

The computational process of the GTCC features is the same as the MFCC extraction scheme. We extracted 14-dim GTCC features from the audio. First, the speech signal is windowed into noticeably short frames of duration 10-50ms to examine over a short period. Making speech signals for a short duration has two purposes i.e., non-stationary signals are assumed to be stationary for such a limited interval of time, and features are extracted efficiently. Subsequently, we employed 48 GT filters to the signal of FFT to compute the energy of the sub-band. Finally, the log of each sub-band is computed followed by applying DCT. The GTCC features are computed as below:

$$GTCC_m = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} \log(X_n) \cos\left[\frac{\pi n}{N}\left(m - \frac{1}{2}\right)\right] 1 \le m \le M(5) \quad (1)$$

Where Xn is the energy of the speech signal in the nth spectral band, N is the number of GT filters and M is the GTCC.

c) Spectral Centroid

Spectral centroid is a measure that is used to characterize the spectrum in digital signal processing. It represents the mid-factor of the mass of the entire energy spectrum, in addition to the power distribution, across the excessive and occasional-frequency bands. The spectral centroid is computed as below:

$$W_e = \frac{\sum_{l=1}^{N} K_e[l] * l}{\sum_{l=1}^{N} K_e[l]} \quad (2)$$

The spectral centroid is computed as below:

Where the Kel is the magnitude of FT at current window e and frequency bin l.

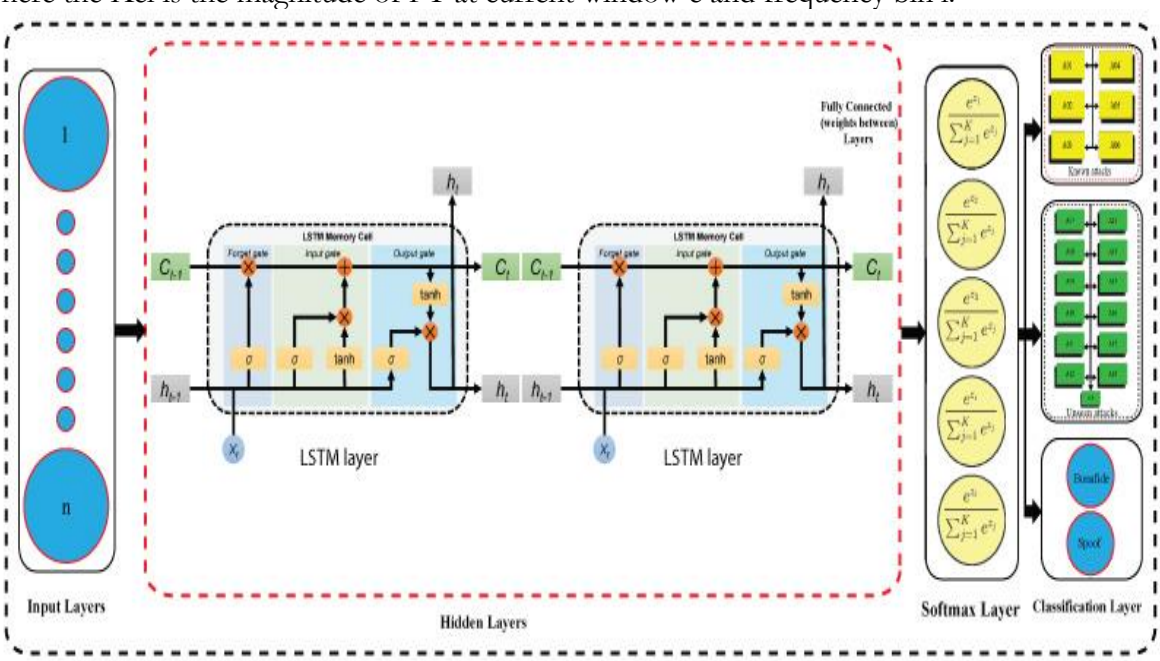

Figure 3. LSTM architecture.

B. Classification

Audio is a time-series data and LSTM is well suited to be used for the classification of time series data. Therefore, in this work, we also employed LSTM for classification purposes. We used different input parameters such as 2, 3, 4, 5, etc., layers, different hidden units such as 100, 200, 300, etc., and different optimizers such as adam, sgd, etc., to get better performance results. After using various configurations, we achieved better classification results on the following parameters: using an adam optimizer, two LSTM layers, 500 hidden units, mini-batch size of 64, and several epochs 25. Figure 3. shows the LSTM architecture being used for all the experimentation purposes.

RESULTS AND DISCUSSION

A. Dataset

We used the ASVspoof2019 LA dataset for experimentation purposes. This dataset consists of two different subset datasets i.e., ASVspoof2019 PA and ASVspoof2019 LA. PA contains samples of voice replay attacks, and the LA dataset contains synthetic and converted speech. Each dataset is further subdivided into three subsets i.e., training, development, and evaluation sets. The details of the ASVspoof2019 LA dataset are given in below Table 1.

B. Evaluation Metrics

To evaluate the performance of the proposed system, we used an accuracy, Equal error rate (EER), F1-score, precision, and recall. Countermeasures having lower EER values indicate better classification performance of the systems to detect spoofing attacks. We compared the performance of our method with baseline methods and other existing systems based on an EER value.

Table 1. Details of the ASVspoof2019 LA dataset.

Subset	Bonafide	Spoof	
	#Utterence	#Utterence	Attacks
Training	2,580	22,800	A01-A06
Development	2,548	22,296	A01-A06
Evaluation	7,355	63,882	A07-A19

C. Performance evaluation on known attacks

The objective of this experiment is to classify the cloning algorithms. Six different TTS and VC cloning algorithms i.e., A01, A02, A03, A04, A05, and A06 are used to generate spoof samples of the ASVspoof2019 LA dataset. These attacks are also called known attacks. There are 22,800 spoof samples of the training set and 22,296 development set that is generated by using these 6 cloning algorithms. Each algorithm-generated 3,800 samples of training and 3,716 of the development set. A01, A02, A03, and A04 are TTS while A05 and A06 are VC algorithms. We used the training samples for training the model and the development set for the testing purpose. From the results reported in Table 2, we can observe that the proposed system successfully classified all the spoofing systems. Moreover, our method performed well on A06 and achieved an accuracy of 99.70%, EER of 0.10%, a precision of 99.70%, recall of 100%, and an F1-score of 99.85%. The system performed second-best on A04 and achieved an accuracy of 99.40%, EER of 0.40%, the precision of 100%, recall of 98.80%, and F1-score of 99.39% while the proposed system performed worst on A03 and achieved an accuracy of 91.30%, EER of 5.93%, the precision of 92.63%, recall of 89.30%, and F1-score of 90.93%. The detailed results of the spoofing systems in terms of accuracy, EER, F1-score, precision, and recall are reported in Table 2. Overall, our system performed well and successfully detected all the cloning algorithms. From the results, we can conclude that the proposed system is robust to capture the variations in signals of spoof audios generated by cloning algorithms.

Table 2. Performance evaluation on the Cloning algorithms/spoofing systems.

Spoofing System	Accuracy%	EER%	F1-score%	Precision%	Recall%
A01	99.10	0.30	99.55	99.10	100
A02	95.60	3.23	95.42	91.24	100
A03	91.30	5.93	90.93	92.63	89.30
A04	99.40	0.40	99.39	100	98.80
A05	91.50	2.73	97.77	92.90	98.83
A06	99.70	0.10	99.85	99.70	100

Performance evaluation on unseen attacks

The objective of this experiment is to evaluate the performance of the proposed system to detect unseen LA attacks i.e., A07, A08, A09, A10, A11, A12, A13, A14, A15, A16, A17, A18, and A19. The evaluation set of the ASVspoof2019 LA dataset contains spoof samples of unseen attack types while the training and development sets contain spoofed samples of known attack types as discussed in section 3.3. We used 25,380 samples of the training set for training the model and 6,000 samples of unseen attacks for testing purposes. From the results reported in Figure4, it is observed that the proposed system performs well on A08 attacks and achieved an accuracy of

96.89%. Our method performed worst on the A17 unseen attack and achieved an accuracy of 80.01%. Overall, the system has reliable performance results on unseen attack types. The detailed results of our method to detect unseen attacks in terms of accuracy are given in Fig 4. From the effects pronounced in determine 4, we will take a look at that the proposed machine has correctly detected all unseen l. a. attacks. Our method is likewise dependable for use for the detection of unseen la spoofing assaults generated through the effective VC, TTS, VC-TTS spoofing algorithms. The proposed system is robust and capable to capture the algorithmic artifacts produced by VC, TTS, VC-TTS algorithms in bonafide audio with better accuracy.

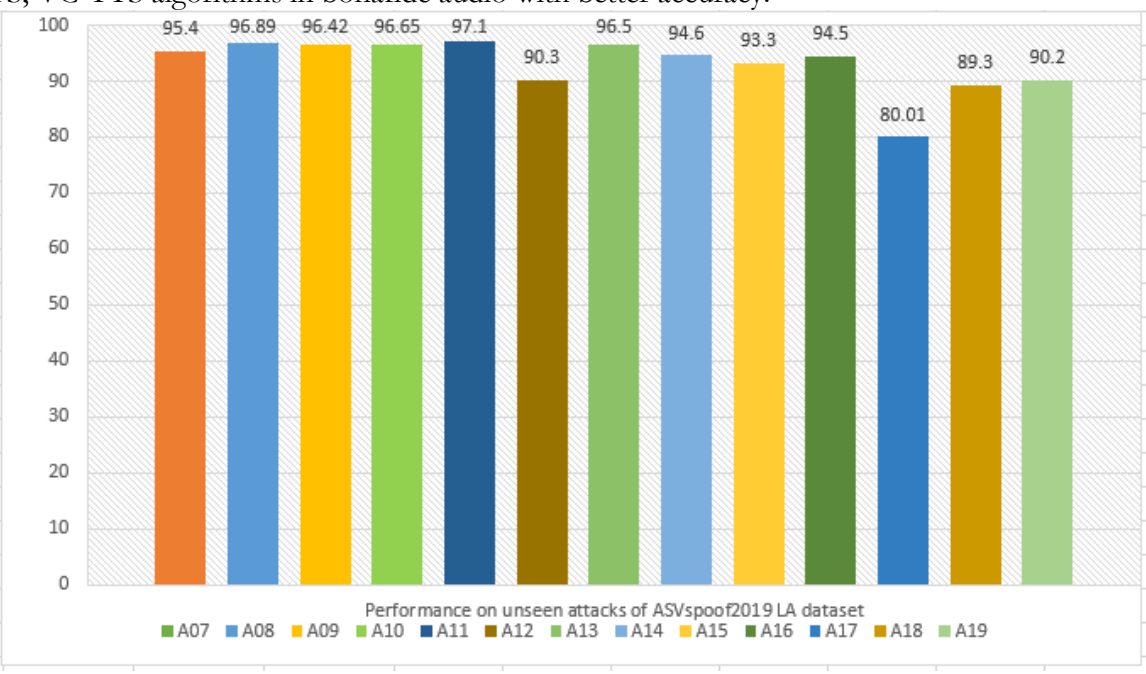

Figure 4. Performance evaluation on unseen attacks.

D. Performance evaluation on LA attacks

We designed this experiment to evaluate the performance of the proposed system to detect LA attacks. For this purpose, we extracted 29-dim spectral features comprising of (14-dim MFCC, 14-dim GTCC, and 1-dim SC) from the ASVSpoof2019 LA dataset. We used 20,005 samples of the training set for training the LSTM model and 18,483 samples of an evaluation set for testing the trained model. Experimental results are shown in Figure 5. The proposed method achieved an accuracy of 98.93%, EER of 1.07%, the precision of 100%, recall of 98.77%, and F1-score of 99.38%. The baseline method (CQCC-GMM [7] and LFCC-GMM [7]) achieved an EER of 9.57% and 8.09%, whereas, the proposed method (MFCC-GTCC-Spectral centroid-LSTM) achieved an 8.5% smaller EER than the baseline methods. From the results reported in Figure5, we can conclude that our method performs well in terms of accuracy and EER. The proposed system is capable of capturing the most discriminatory characteristics from audio signals of spoofed samples generated by TTS and VC algorithms. Experimental results signify the effectiveness of our method that can be implemented in ASV systems for the reliable detection of LA attacks.

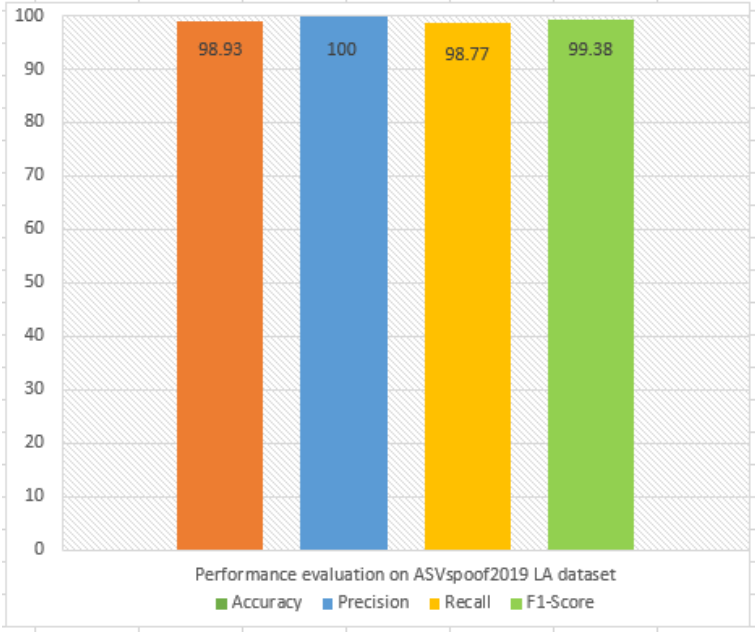

Figure 5. Performance evaluation on ASVspoof2019 LA dataset.

E. Comparison with other methods

We designed this experiment to check the effectiveness of the proposed system against other existing state-of-the-art methods [22, 14, 16, 18, 21, 17, 19, 15]. As shown in Table 3, we compared the results of our method with the baseline [7], and other existing state-of-the-art

methods to demonstrate the superiority of the proposed system. The EER and mint DCF values of the existing state-of-the-art methods and our system are reported in Table 3. It can be observed that our method performed well against all the existing methods and baselines methods [7]. Chen et al. [19], perform the second-best by achieving an EER of 3.49% and min-tDCF of 0.092, whereas, the baseline methods (CQCC-GMM and LCFF-GMM) [7] performed worst by achieving an EER of 9.57% and 8.09%. From the results reported in Table 3, we observe an accuracy gain of 8.5%, 7.02%, 6.59%, 5.31%, 5.21%, and 4.52% than the state-of-the-art methods [7,26,17,20,22]. These results signify the effectiveness and superiority of the proposed spoofing countermeasure. Experimental results and comparative analysis show that our voice anti-spoofing technique outperforms all the existing techniques in terms of EER and min-tDCF values. From these consequences, we can finish that the proposed machine can reliably be used for the detection of la attacks.

Table 3. Performance comparison with other methods.

System	EER%	min-tDCF
Baseline (CQCC-GMM) [7]	9.57	0.237
Baseline (LFCC-GMM) [7]	8.09	0.212
Chettri et al. [26]	7.66	0.179
Monterio et al. [17]	6.38	0.142
Gomez-Alanis et al. [20]	6.28	--
Aravind et al. [22]	5.32	0.151
Lavrentyeva et al. [25]	4.53	0.103
Wu et al. [21]	4.07	0.102
Tak et al. [23]	3.50	0.090
Chen et al. [19]	3.49	0.092
Proposed (MFCC-GTCC-Spectral Centroid-LSTM)	1.07	0.0343

CONCLUSION

This paper has presented a novel voice spoofing countermeasure to locate l. a. assaults. additionally, the proposed gadget is capable to come across the spoofing systems (TTS and VC) that have been used to generate the spoofed samples of the ASVspoof2019 l. a. dataset. We proposed a novel set of integrated features, which captures maximum alterations and algorithmic artifacts present in speech signals. We employed LSTM for classification purposes to discriminate authentic and fake audio, classify the known attack types, and detect all the unseen LA attacks. Experimental results show that the proposed system outperformed the baseline and the existing state-of-the-art methods. The proposed system gives better classification results by achieving an 8.5% smaller EER than the baseline [7]. In the future, we aim to apply the proposed system on PA and deepfake datasets.

References

[1] Voiceprint: The New WeChat Password. 2015 [cited 2021 04 April]; Available from: https://blog.wechat.com/2015/05/21/voiceprint-the-new-wechat-password/.

[2] Millward, S. Open Sesame: Baidu Helps Lenovo Use Voice Recognition to Unlock Android Phones. 2012 [cited 2020 12 March]; Available from: https://www.techinasia.com/baidu-lenovo-voice-recognition-android-unlock.

[3] Vigderman, A. What Is Home Automation and How Does It Work? Mar 15, 2021 [cited 2021 April 16]; Available from: https://www.security.org/home-automation/.

[4] Fernández, L. Efma recognizes Garanti Bank's mobile voice assistant. 2017 [cited 2020 3 March]; Available from: https://www.bbva.com/en/efma-recognizes-garanti-banks-mobile-voice-assistant/.

[5] K. Delac and M. Grgic, "A survey of biometric recognition methods," in Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine. IEEE, 2004, pp. 184–193.

[6] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification: A survey," speech communication, vol. 66, pp. 130–153, 2015.

[7] M. Todisco, X. Wang, V. Vestman, M. Sahidullah, H. Delgado, A. Nautsch, J. Yamagishi, N. Evans, T. H. Kinnunen, and K. A. Lee, "ASVspoof 2019: Future horizons in spoofed and fake audio detection," Proc. Interspeech, pp. 1008–1012, 2019.

[8] M. R. Kamble, H. B. Sailor, H. A. Patil, and H. Li, "Advances in anti-spoofing: from the perspective of ASVspoof challenges," APSIPA Transactions on Signal and Information Processing, vol. 9, 2020.

[9] R. K. Das, T. Kinnunen, W.-C. Huang, Z.-H. Ling, J. Yamagishi, Z. Yi, X. Tian, and T. Toda, "Predictions of subjective ratings and spoofing assessments of voice conversion challenge 2020 submissions," in Proc. Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020, 2020, pp. 99–120.

[10] Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilc¸i, M. Sahidullah, and A. Sizov, "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge," in Sixteenth Annual Conference of the International Speech Communication Association, 2015, Conference Proceedings.

[11] T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi, and K. A. Lee, "The ASVspoof 2017 challenge: Assessing the limits of replay spoofing attack detection," in Proc. Interspeech 2017, 2017, pp. 2–6. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2017-1111

[12] T. B. Patel and H. A. Patil, "Combining evidence from Mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech," in Sixteenth Annual Conference of the International Speech Communication Association, Conference Proceedings.

[13] M. Sahidullah, T. Kinnunen, and C. Hanilc¸i, "A comparison of features for synthetic speech detection," in Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[14] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in Proc. Odyssey, vol. 45, 2016, pp. 283–290. [Online]. Available: http://dx.doi.org/10.21437/Odyssey.2016-41

[15] L. Wang, Y. Yoshida, Y. Kawakami, and S. Nakagawa, "Relative phase information for detecting human speech and spoofed speech," in Sixteenth Annual Conference of the International Speech Communication Association, Conference Proceedings.

[16] J. Sanchez, I. Saratxaga, I. Hernaez, E. Navas, D. Erro, and T. Raitio, "Toward a universal synthetic speech spoofing detection using phase information," IEEE Transactions on Information Forensics and Security, vol. 10, no. 4, pp. 810–820, 2015.

[17] C. Zhang, C. Yu, and J. H. Hansen, "An investigation of deep-learning frameworks for speaker verification antispoofing," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 4, pp. 684–694, 2017.

[18] J. Monteiro, J. Alam, and T. H. Falk, "Generalized end-to-end detection of spoofing attacks to automatic speaker recognizers," Computer Speech & Language, p. 101096, 2020.

[19] T. Chen, A. Kumar, P. Nagarsheth, G. Sivaraman, and E. Khoury, "Generalization of audio deepfake detection," in Proc. Odyssey the Speaker and Language Recognition Workshop, 2020, Conference Proceedings, pp. 132–137.

[20] A. Gomez-Alanis, A. M. Peinado, J. A. Gonzalez, and A. M. Gomez, "A light convolutional GRU-RNN deep feature extractor for ASV spoofing detection," Proc. Interspeech, pp. 1068–1072, 2019.

[21] Z. Wu, R. K. Das, J. Yang, and H. Li, "Light convolutional neural network with feature genuinization for detection of synthetic speech attacks," Proc. Interspeech, pp. 1101–1105, 2020.

[22] P. Aravind, U. Nechiyil, N. Paramparambath, et al., "Audio spoofing verification using deep convolutional neural networks by transfer learning," arXiv preprint arXiv:2008.03464, 2020.

[23] H. Tak, J. Patino, A. Nautsch, N. Evans, and M. Todisco, "Spoofing attack detection using the non-linear fusion of sub-band classifiers," Proc. Interspeech, pp. 1106–1110, 2020.

[24] X. Tian, Z. Wu, X. Xiao, E. S. Chng, and H. Li, "Spoofing detection from a feature representation perspective," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 2119–2123.

[25] G. Lavrentyeva, S. Novoselov, A. Tseren, M. Volkova, A. Gorlanov, and A. Kozlov, "STC antispoofing systems for the ASVspoof2019 challenge," Proc. Interspeech, pp. 1033–1037, 2019.

[26] B. Chettri, D. Stoller, V. Morfi, M. A. M. Ram´ırez, E. Benetos, and B. L. Sturm, "Ensemble models for spoofing detection in automatic speaker verification," in Proc. Interspeech, 2019, pp. 1018–1022. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2505

[27] Alzantot, M., Z. Wang, and M.B. Srivastava, Deep residual neural networks for audio spoofing detection. arXiv preprint arXiv:1907.00501, 2019.

[28] Lai, C.-I., et al., ASSERT: Anti-spoofing with squeeze-excitation and residual networks. arXiv preprint arXiv:1904.01120, 2019.

[29] Gomez-Alanis, A., et al., A Light Convolutional GRU-RNN Deep Feature Extractor for ASV Spoofing Detection. Proc. Interspeech 2019, 2019: p. 1068-1072.

[30] Alam, M.J., et al. Spoofing Detection on the ASVspoof2015 Challenge Corpus Employing Deep Neural Networks. in Odyssey. 2016.

[31] Qian, Y., N. Chen, and K. Yu, Deep features for automatic spoofing detection.Speech Communication, 2016. 85: p. 43-52.

[32] Access on 9.10.2021, Available online at: http://cobramoto.fi/phlms/growth-of-voice-assistants

[33] Access on 9.10.2021, Available online at: lhttps://venturebeat.com/2021/07/14/how-voice-biometrics-is-saving-financial-services-companies-millions-and-eliminating-fraud/.