Multi-band Spectral Entropy Information for Detection of Replay Attacks

Yitong Liu, Rohan Kumar Das and Haizhou Li
National University of Singapore, Singapore
E-mail: e0008049@u.nus.edu, {rohankd, haizhou.li}@nus.edu.sg

***Abstract*—Replay attacks have been proven to be a potential threat to practical automatic speaker verification systems. In this work, we explore a novel feature based on spectral entropy for the detection of replay attacks. The spectral entropy is a measure to capture spectral distortions and flatness. It is found that the replay speech carries artifacts in the process of recording and playback. We hypothesize that spectral entropy can be a useful information to capture such artifacts. In this regard, we explore multi-band spectral entropy feature for replay attack detection. The studies are conducted on ASVspoof 2017 Version 2.0 database that deals with replay speech attacks. A baseline system with popular constant-Q cepstral coefficient (CQCC) feature is also developed. Finally, a combined system is proposed with multi-band spectral entropy and CQCC features that outperforms the baseline. The experiments validate the idea of multi-band spectral entropy feature.**

I. Introduction

Automatic speaker verification (ASV) deals with authenticating a claimed identity for a given speech [1]–[3]. It has increased attention from the community due to various possible applications in the recent years. Some of the applications already have shown success for practical systems [4]–[9]. However, such systems under practice are found to be vulnerable to the spoofing attacks [10], [11]. In general, there are four broad categories of spoofing attacks, namely, impersonation, text-to-speech, voice conversion and replay attacks [10], [11]. Among these text-to-speech and voice conversion are synthetic speech attacks, whereas impersonation is a behavioral attack. Again, replay attacks, also known as presentation attacks are one of the most easiest way to perform spoofing attacks by using recorded speech samples of a particular speaker. In this work, we focus on replay attacks that are the most common way for spoofing speaker identity. Figure 1 shows an illustration of replay attacks, where a replayed speech signal is used for unauthorized access.

The countermeasures to spoofing attacks are designed to identify the spoofing attacks based on the artifacts extracted from given speech. In this regard, a myriad of front-end methods have been explored that can capture relevant information to identity such attacks. The earlier works focused on far-field recording based replay attacks that used noise and reverberation to classify spoofed speech [12], [13]. The authors of [14] proposed a spectral bitmap based method to identify the replay attacks for text-dependent speaker verification. Later, such methods are applied to replay detection in text-independent speaker verification in terms of average spectral bitmap models [15]. Another study based on spectral features and score normalization was carried out for replay speech detection in [16].

Fig. 1. An illustration of replay spoofing attacks.

The drive to spearhead anti-spoofing research in terms of ASVspoof series of challenge led many recent works to design various countermeasures [17]. In the year 2017, ASVspoof 2017 challenge was organized that focuses particularly on replay speech detection [18]. The constant-Q cepstral coefficient (CQCC) with Gaussian mixture model (GMM) forms the baseline for the challenge [19], [20]. It is to be noted that CQCC feature has been found to be very effective for synthetic speech detection and replay speech is also no exception to it compared to existing features. A comparative analysis of different features for replay detection is presented in [21]. Few other explorations by various participating groups in that challenge include epoch strength and peak to side lobe ratio based features [22], variable length Teager energy separation based instantaneous frequency feature [23], high frequency features [24], phase features [25] and hierarchical scattering decomposition coefficients [26]. The studies reported in these works show the importance of front-end features for detection of replay attacks.

During post ASVspoof 2017 challenge evaluation, few anomalies are found out that showed beep sounds and few broken examples. The presence of beep sounds and broken examples has a definite impact on the overall performance. Therefore, those beep sounds were cut out along with the removal of broken examples and a second version 2.0 of the database was released [27]. A comparison of GMM and i-vector systems is reported that showed the former is more useful for modeling in this task [28], [29]. Further, the usefulness of log-energy feature as well as cepstral mean and variance normalization (CMVN) [30] has been shown to contribute towards improved detection of spoofing attacks. Followed by the release of ASVspoof 2017 Version 2.0 database, several

investigations have been made to detect replay attacks. Some of these include extended CQCC features [31], modulation spectrum [32], instantaneous phase, excitation source features [33] and low frequency frame normalization [34]. Again, the importance of various long range acoustic features and deep features are shown recently in [35], [36].

The investigations on various countermeasures show that the front-end features are more effective for detection of spoofing attacks. Therefore, we focus on finding novel front-end countermeasure in this work. Most of these existing front-end features capture either magnitude or phase information of the signal with some representation. The replay speech is generated using recorder and a playback device, thus the produced replayed signal possesses the device characteristics as well as the background environment information. The effects thus incurred in the replayed signal have definite impact on the spectrum that can be used as artifacts for discriminating it from genuine speech.

The entropy is originally defined for information sources by Shannon [37]. It plays a central role in the context of pattern classification and information technology as a metric of disorganization or uncertainty in a random variable, such as information and choice. The concept of entropy has been adopted to the field of speech technology to apply in voice activity detection [38] and speech recognition [39], exhibiting significantly improved performance. In this work, we explore a multi-band spectral entropy feature that measures the power spectral flatness of the spectrum to capture the distortions present in replayed speech. The use of multi-band is motivated by effectiveness of various subbands features for spoof detection [40]. The proposed feature is used to extract the formants variations and their temporal locations in the spectrum, which is observed to be different for genuine and replay speech. The studies in this work are conducted on ASVspoof 2017 Version 2.0 corpus for replay speech detection.

The rest of the paper is organized as follows. Section II details the multi-band spectral entropy feature. The results and experiments are reported in Section III. Finally, Section IV concludes the work.

II. Multi-band Spectral Entropy

The spectral entropy is a measure to capture the spectral distortions and spectral flatness. It is computed over short-term processed speech signal. For every frame of the signal, the power spectral density is computed using fast Fourier transform (FFT). The raw spectrum also contains pitch information and therefore it is smoothed by applying a linear filter bank before entropy computation to emphasize the high frequency component. To compute the entropy of a spectrum, the spectrum should be like a probability mass function (PMF), where the area under the spectrum sums up to 1. Therefore, the individual frequency components of the short-term Fourier transform (STFT) spectrum is divided by the sum of all the components to convert the spectrum into a PMF like function.

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i}, \text{for } i = 1 \text{ to } N \tag{1}$$

Algorithm 1 : Multi-band Spectral Entropy

1: Let $x(n)$ be the input speech signal.
2: Pre-process $x(n)$ by a pre-emphasis filter to obtain $x_p(n)$.
3: For each short term frame of 20 ms, i.e., x_f, multiply it with Hamming window to obtain x_h.
4: Compute FFT for the windowed signal x_h to obtain $X(w)$.
5: Compute the power spectral density (PSD) from the FFT, i.e., $X(w)$ to obtain PSD spectrum

$$PSD = \frac{|X(w)|^2}{N}$$

6: Multiply PSD with the linear filter bank to obtain PSD_l.
7: For each frequency subband of 200 Hz, convert the m-th subband spectrum to a PMF like function

$$psd_i = \frac{PSD_{l\_i}}{\sum_{i=1}^{N} PSD_{l\_i}},$$

8: Compute the multi-band spectral entropy $H(M)$, where $M = 1, \ldots, m$, for every subband from the normalized spectrum psd_i,

$$H(M) = -\sum_{i=1}^{N} psd_i.\ \log_2(psd_i)$$

where X_i is the power spectral density of i-th frequency component of the spectrum, $\mathbf{x} = x_1, \ldots, x_N$ is the PMF of the spectrum and N is the number of points in the spectrum. The normalized spectrum can be treated as a PMF, where the area under the normalized spectrum sums to 1. By converting a spectrum to PMF, the peak capturing property of entropy can be explored. A PMF with sharp peaks will have low entropy, while a PMF with flat distribution will have high entropy. For each frame, the entropy H is computed from $\mathbf{x}$ as follows [39],

$$H = -\sum_{i=1}^{N} x_i.\ \log_2 x_i \tag{2}$$

The explorations on using entropy from the full-band spectrum is found to be not that effective as it captures only the gross peakiness of the spectrum and cannot resolve the locations for the peaks. Therefore, in order to capture the location of the peaks, the idea of multi-band entropy features was introduced in [41]. The full-band spectrum is divided into several non-overlapping subbands of equal size. Each subband spectrum is then converted into a PMF so that the area under each normalized subband spectrum sums up to 1. Using Equations (1) and (2), the entropy for each subband PMF is separately computed and one entropy value is obtained for each subband. These subband entropy values indicate the presence or absence of peaks in that particular subband. Further, the number of subbands determines the dimensionality of the feature vector. We also note that different components of the entropy feature vector have different dynamic ranges and activation points depending upon occurrence of a peak in

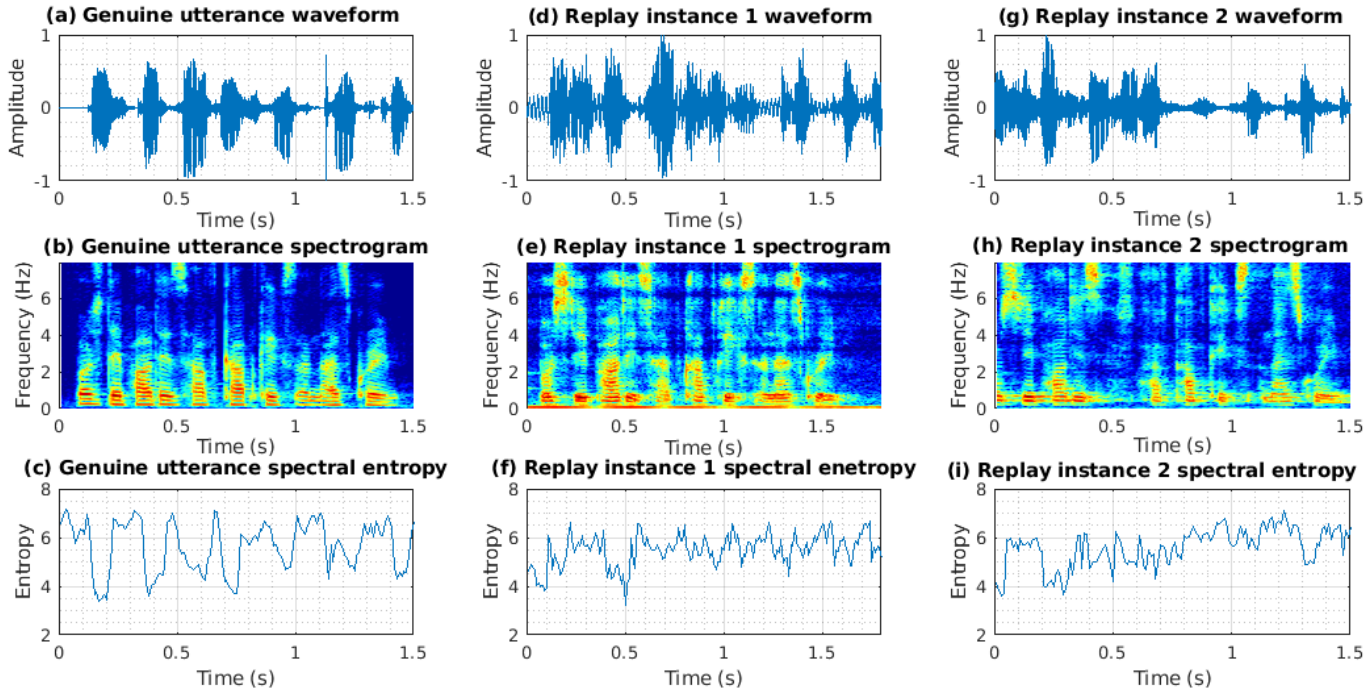

Fig. 2. (a) is waveform of a genuine utterance of "Birthday parties have cupcakes and ice cream", (b) and (c) are corresponding spectrogram and entropy contour. (d) and (g) are waveform of two replayed instances of (a); (e) and (f), (h) and (i) are their corresponding spectrogram and entropy contour, respectively.

a particular subband. Algorithm 1 presents the detailed steps followed to obtain the multi-band spectral entropy feature.

Figure 2 shows the comparison of a genuine speech with its two replayed signal instances taken from ASVspoof 2017 Version 2.0 corpus. On comparing their spectrograms, we observe apparent differences between the genuine and replayed versions. The replayed speech is produced by use of recording and playback device in various environments. Thus, the device and environment characteristics have an impact on the replayed signal. Due to such effects the formant distributions and positions are different in replayed signal from that in the genuine speech signal. It is observed from Figure 2 that there are high frequency contents in the replayed speech instances. These information can play a crucial role as artifacts for detection of replay spoofing attacks. We then observe the entropy contours for the genuine and the replay speech.

Figure 2 (c), (f) and (i) show that the spectral entropy for genuine speech signal has smooth pattern and manages to track most of the formants which are represented by the contour unlike the case with replay speech. For the genuine spectrogram with distinct and clear peaks, the voiced regions of a speech signal induce low entropy and the unvoiced regions produce higher entropy, which results in a high variance of entropy. On the contrary, for the replayed speech signal with noise region introduced by environment, recording and playback devices, the spectrogram has a flatter and more even distribution and the corresponding entropy is obtained with modest amplitude on average. The variance of entropy for the spoofed utterance is comparatively smaller than that of the corresponding genuine speech. This is because the entropy obtained from the voiced and noisy regions do not differ much for the replayed utterance compared to that in genuine speech. Thus, the entropy of a signal can capture useful information that be effective for detection of replay attacks.

III. Experimental Results and Analysis

In this section, we discuss the details of the database, experimental setup and the results of various studies conducted for replay speech detection.

A. Database

The ASVspoof 2017 database is constituted by a subset of RedDots data [42] collection and its replay derivatives [27]. The genuine speech is obtained directly from the original corpus and replay attack based speech utterances are replayed versions of the speech from RedDots corpus. The RedDots corpus is designed for text-dependent speaker verification and therefore each utterance has around 2-3 seconds duration. The replay derivatives are generated using various recording and playback devices in different environments. The dataset is designed to contain a diverse range of replay configurations with a unique combination of recording environment, replay and recording device ranging from conditions. The dataset has three non-overlapping subsets, namely, training, development and evaluation set. The train and development set are used to build models and optimize the parameters for the studies. The studies are then to be repeated on the evaluation set for reporting results. Further, there are less number of variabilities in terms of recording configurations in train and development set. On the other hand, the evaluation set has many unseen devices and replay is made in different recording scenarios. Equal error rate (EER) is used as a metric to report the

TABLE I
SUMMARY OF ASVSPOOF 2017 VERSION 2.0 CORPUS [27].

Database Subset	# Speakers	# Replay Configurations	# Utterances	
			Genuine	Spoofed
Train	10	3	1,507	1,507
Development	8	10	760	950
Evaluation	24	57	1,298	12,008

performance for replay attack detection. Table I shows the detailed composition of ASVspoof Version 2.0 corpus.

B. Experimental Setup

A baseline system is implemented before considering the system with multi-band spectral entropy feature. The CQCC features with log-energy coefficients are used for the baseline system. The CQCC features are derived using constant-Q transform (CQT) based long-term window transform. It is to be noted that we followed the same configurations while extracting CQCC features as mentioned by the original authors [19], [20]. The studies in [27] showed that GMM based system outperformed the results of i-vector based system. Hence, we have chosen GMM as the test bed for our models. Two GMMs of 512 mixture components are then learned to have models for genuine and replay speech. We note that the examples of the train set are used to build these models that are evaluated on the development set. However, we combined the train and development set examples to learn the models for the studies with evaluation set. Given a test speech, its CQCC features are extracted then log-likelihood is computed with respect to both the models to obtain a log-likelihood ratio. The score thus obtained is compared to a threshold for detecting the replay attacks.

In case of multi-band spectral entropy feature, a pre-emphasis filter is first applied to the speech signal to exploit high frequency information [43]. The speech signal is short-term processed with 20 ms Hamming window with a shift of 10 ms. We have varied the number of subbands and have fixed as 40 according to the best possible results on the development set. Thus, the static feature dimension of the multi-band spectral entropy feature is 40. The delta (Δ) and delta delta ($\Delta\Delta$) features are also extracted for the studies. On top of the extracted features, CMVN is applied to fit it to zero mean and unit variance. Once the features are extracted GMM models are build as explained in the case of baseline system and rest of the system pipeline remains the same.

In this work, we also study the complementary information carried by the spectral entropy feature to that carried by the CQCC features. A score level combination of the results of the two systems are made as given in [44], [45]. We note that the weights of the two systems are learned on the development set and then applied on the evaluation set. The fusion system is expected to provide improved results over the baseline due to the different nature of information carried by each feature.

C. Results and Discussion

Table II shows the performance of multi-band spectral entropy feature on ASVspoof Version 2.0 corpus. The feature is studied for different coefficient configurations. It is found that the configuration Δ gives the optimal result on the development set. We have therefore highlighted the results under that setting. This kind of trend to have better results for dynamic coefficients is well supported by previous investigations for different features [19], [20]. Then the CQCC feature based well known system is also evaluated that forms the baseline. Further, we consider another contrast system based on mel frequency cepstral coefficient (MFCC) feature for comparison [46]. Table III reports the results of these three features for comparison. We observe that the multi-band spectral entropy feature based system outperforms the system with MFCC features showing importance of entropy information than the traditional way of capturing spectral information. The CQCC based system performs better than the two systems as it contains the long-term signal characteristics.

TABLE II
PERFORMANCE IN EER (%) OF MULTI-BAND SPECTRAL ENTROPY FEATURE FOR DIFFERENT COEFFICIENT CONFIGURATIONS.

Configuration	Development	Evaluation
S	13.55	18.06
Δ	**13.30**	**17.70**
$\Delta\Delta$	15.70	19.61
$S + \Delta$	13.50	17.52
$\Delta + \Delta\Delta$	14.52	18.43
$S + \Delta + \Delta\Delta$	13.25	18.85

TABLE III
PERFORMANCE IN EER (%) OF FUSION SYSTEM AND COMPARISON.

System	Development	Evaluation
Multi-band Spectral Entropy	13.30	17.70
MFCC	18.04	20.78
CQCC	8.93	12.64
Multi-band Spectral Entropy + CQCC Fusion	**7.33**	**11.16**

We then perform the score level fusion of systems with multi-band spectral entropy and CQCC features. The fusion of the two systems is tuned on the development set. Table III shows that the fused system performs the best among three different systems considered. This depicts the complementary nature of information captured by spectral entropy from widely popular CQCC features that is useful for replay attack detection. Further, it confirms the complementary nature of information captured by both the features. The future work will focus on using spectral entropy with long-term features to have improved detection of replay attacks.

IV. CONCLUSION

This work focuses on exploring spectral entropy extracted from the subbands as a novel attribute for replay speech detection. The spectral entropy captures the spectral distortions and flatness that are unique for genuine and replay speech. The multi-band spectral entropy feature extracted is used to study the replay attacks using ASVspoof 2017 Version 2.0 database. The studies reveal that the spectral entropy has

definite characteristics to distinguish replay speech from the genuine counterparts. Further, the fusion of such information with long-term CQCC features shows fruitful results highlighting its importance for detection of replay attacks.

V. Acknowledgements

This research is supported by Programmatic Grant No. A1687b0033 from the Singapore Government's Research, Innovation and Enterprise 2020 plan (Advanced Manufacturing and Engineering domain), and as the result of a research collaboration between the Human Language Technology Lab at NUS and Kuai Shang Tong Technology Co. Ltd. The work was carried out as part of the final year project of Yitong Liu during her undergraduate study.

References

[1] J. P. Campbell, "Speaker recognition: a tutorial," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1437–1462, 1997.

[2] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech communication*, vol. 52, no. 1, pp. 12–40, 2010.

[3] J. H. L. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74–99, Nov 2015.

[4] K.-A. Lee, B. Ma, and H. Li, "Speaker verification makes its debut in smartphone," in *SLTC Newsletter*, February 2013.

[5] K.-A. Lee, A. Larcher, H. Thai, B. Ma, and H. Li, "Joint application of speech and speaker recognition for automation and security in smart home," in *INTERSPEECH*, 2011, pp. 3317–3318.

[6] D. Chakrabarty, S. R. M. Prasanna, and R. K. Das, "Development and evaluation of online text-independent speaker verification system for remote person authentication," *International Journal of Speech Technology*, vol. 16, no. 1, pp. 75–88, 2013.

[7] S. Dey, S. Barman, R. K. Bhukya, R. K. Das, Haris B C, S. R. M. Prasanna, and R. Sinha, "Speech biometric based attendance system," in *National Conference on Communications (NCC) 2014, IIT Kanpur*, 2014.

[8] R. K. Das, S. Jelil, and S. R. M. Prasanna, "Development of multi-level speech based person authentication system," *Journal of Signal Processing Systems*, vol. 88, no. 3, pp. 259–271, Sep 2017.

[9] S. Jelil, A. Shrivastava, R. K. Das, S. R. M. Prasanna, and R. Sinha, "SpeechMarker: a voice based multi-level attendance application," in *Interspeech 2019*, 2019.

[10] 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.

[11] M. Sahidullah, H. Delgado, M. Todisco, T. Kinnunen, N. Evans, J. Yamagishi, and K.-A. Lee, *Introduction to Voice Presentation Attack Detection and Recent Advances*. Cham: Springer International Publishing, 2019, pp. 321–361.

[12] J. Villalba and E. Lleida, "Preventing replay attacks on speaker verification systems," in *International Carnahan Conference on Security Technology (ICCST) 2011*, pp. 1–8.

[13] ——, "Detecting replay attacks from far-field recordings on speaker verification systems," in *Biometrics and ID Management*. Springer, 2011, pp. 274–285.

[14] Z. Wu, S. Gao, E. S. Chng, and H. Li, "A study on replay attack and anti-spoofing for text-dependent speaker verification," in *Proc. Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2014.

[15] A. Paul, R. K. Das, R. Sinha, and S. R. M. Prasanna, "Countermeasure to handle replay attacks in practical speaker verification systems," in *International Conference on Signal Processing and Communications (SPCOM) 2016*, June 2016, pp. 1–5.

[16] J. Gaka, M. Grzywacz, and R. Samborski, "Playback attack detection for text-dependent speaker verification over telephone channels," *Speech Communication*, vol. 67, pp. 143 – 153, 2015.

[17] Z. Wu, J. Yamagishi, T. Kinnunen, C. Hanili, M. Sahidullah, A. Sizov, N. Evans, M. Todisco, and H. Delgado, "ASVspoof: The automatic speaker verification spoofing and countermeasures challenge," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 588–604, June 2017.

[18] 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.

[19] M. Todisco, H. Delgado, and N. Evans, "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," *Computer Speech & Language*, vol. 45, pp. 516–535, 2017.

[20] ——, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in *Odyssey 2016*, 2016, pp. 283–290.

[21] R. Font, J. M. Espn, and M. J. Cano, "Experimental analysis of features for replay attack detection results on the ASVspoof 2017 challenge," in *Proc. Interspeech 2017*, 2017, pp. 7–11.

[22] S. Jelil, R. K. Das, S. R. M. Prasanna, and R. Sinha, "Spoof detection using source, instantaneous frequency and cepstral features," in *Proc. Interspeech 2017*, 2017, pp. 22–26.

[23] H. A. Patil, M. R. Kamble, T. B. Patel, and M. H. Soni, "Novel variable length teager energy separation based instantaneous frequency features for replay detection," in *Proc. Interspeech 2017*, 2017, pp. 12–16.

[24] M. Witkowski, S. Kacprzak, P. elasko, K. Kowalczyk, and J. Gaka, "Audio replay attack detection using high-frequency features," in *Proc. Interspeech 2017*, 2017, pp. 27–31.

[25] K. Srinivas, R. K. Das, and H. A. Patil, "Combining phase-based features for replay spoof detection system," in *International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Taipei, Taiwan, 2018, pp. 151–155.

[26] K. Sriskandaraja, G. Suthokumar, V. Sethu, and E. Ambikairajah, "Investigating the use of scattering coefficients for replay attack detection," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) 2017*, Dec 2017, pp. 1195–1198.

[27] H. Delgado, M. Todisco, M. Sahidullah, N. Evans, T. Kinnunen, K. A. Lee, and J. Yamagishi, "ASVspoof 2017 Version 2.0: meta-data analysis and baseline enhancements," in *ODYSSEY 2018, The Speaker and Language Recognition Workshop, June 26-29, 2018, Les Sables d'Olonne, France*, Les Sables d'Olonne, France, 2018.

[28] D. A. Reynolds and R. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, Jan 1995.

[29] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, May 2011.

[30] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 254–272, Apr 1981.

[31] J. Yang, R. K. Das, and H. Li, "Extended constant-Q cepstral coefficients for detection of spoofing attacks," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honolulu, Hawaii, 2018, pp. 1024–1029.

[32] G. Suthokumar, V. Sethu, C. Wijenayake, and E. Ambikairajah, "Modulation dynamic features for the detection of replay attacks," in *Interspeech*, 2018, pp. 691–695.

[33] R. K. Das and H. Li, "Instantaneous phase and excitation source features for detection of replay attacks," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honolulu, Hawaii, 2018, pp. 1030–1037.

[34] J. Yang and R. K. Das, "Low frequency frame-wise normalization over constant-q transform for playback speech detection," *Digital Signal Processing*, vol. 89, pp. 30 – 39, 2019.

[35] R. K. Das, J. Yang, and H. Li, "Long range acoustic features for spoofed speech detection," in *Interspeech 2019*, 2019.

[36] ——, "Long range acoustic and deep features perspective on ASVspoof 2019," in *Automatic Speech Recognition Understanding Workshop (ASRU) 2019 (Submitted)*, 2019.

[37] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[38] P. Renevey and A. Drygajlo, "Entropy based voice activity detection in very noisy conditions," in *Seventh European Conference on Speech Communication and Technology*, 2001.

[39] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2004*, vol. 1, 2004, pp. I–193.

[40] J. Yang, R. K. Das, and H. Li, "Significance of subband features for synthetic speech decetion," *IEEE Transactions on Information Forensics and Security (Accepted with Minor Revision)*, 2019.

[41] H. Misra, S. Ikbal, S. Sivadas, and H. Bourlard, "Multi-resolution spectral entropy feature for robust asr," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2005.*, vol. 1, 2005, pp. I–253.

[42] T. Kinnunen, M. Sahidullah, M. Falcone, L. Costantini, R. G. Hautamäki, D. Thomsen, A. Sarkar, Z.-H. Tan, H. Delgado, M. Todisco *et al.*, "Reddots replayed: A new replay spoofing attack corpus for text-dependent speaker verification research," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2017*, 2017, pp. 5395–5399.

[43] R. Vergin and D. O'Shaughnessy, "Pre-emphasis and speech recognition," in *Canadian Conference on Electrical and Computer Engineering*, vol. 2, 1995, pp. 1062–1065.

[44] R. K. Das, Abhiram B., S. R. M. Prasanna, and A. G. Ramakrishnan, "Combining source and system information for limited data speaker verification," in *Interspeech 2014, Singapore*, 2014, pp. 1836–1840.

[45] R. K. Das and S. R. M. Prasanna, "Exploring different attributes of source information for speaker verification with limited test data," *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 184–190, 2016.

[46] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug 1980.