Odyssey 2018 The Speaker and Language Recognition Workshop
26-29 June 2018, Les Sables d'Olonne, France

BOOSTING THE PERFORMANCE OF SPOOFING DETECTION SYSTEMS ON REPLAY ATTACKS USING Q-LOGARITHM DOMAIN FEATURES NORMALIZATION

Jahangir Alam, Gautam Bhattacharya, Patrick Kenny

Computer Research Institute of Montreal (CRIM)
Montreal, Quebec, Canada

{Jahangir.alam, Gautam.bhattacharya}@crim.ca

ABSTRACT

Feature normalization strategies help to compensate for the effects of environmental mismatch and are normally incorporated into the feature extraction framework after applying a logarithmic or power function nonlinearity. For spoofing detection systems in the presence of voice conversion and speech synthesis-based spoofing attacks, feature normalization is found to be harmful. However, when it comes to spoofing detection for replay attacks, normalization of features aids to reduce equal error rates significantly. In this work, we use discrete Fourier transform (DFT)-based spectral and product spectral features with feature normalization applied in the q-log domain. The q-log function acts as intermediate domain between linear and log domains for normalization of the features. After that, the final features are extracted by applying a principal component analysis technique to the log DFT and product power spectra. Experimental results on the version 2 of second ASVspoof2017 challenge evaluation data show that normalizing features in q-log domain results in relative reduction of equal error rates by approximately 5%. Over all four baseline systems, the DFT spectral features, normalized in the q-log domain, provides an average relative improvement of 28%.

1. INTRODUCTION

Speaker recognition researchers acknowledge that systems which aim to verify speakers automatically based on their pronunciation of an utterance are vulnerable to spoofing attacks. A spoofing attack occurs when a person or computer program pretends to be a legitimate user of an authentication system by falsifying data and thereby gaining illegitimate access and advantages. The most widely known spoofing attacks for speaker recognition systems are impersonation, speech synthesis, voice conversion and replay attacks. Impersonation attack requires a mimic to imitate a target speaker's voice to get access to the system. The success of this attack strongly depends on the mimic and the target. Impersonation does not pose a genuine threat as this attack can be detected by reliable speaker verification systems [1-5].

In a speech synthesis attack the fraudster creates a synthetic voice of a target speaker by adapting the speech synthesizer to the target speaker and then plays it to the authentication system to get illegitimate access. In a voice conversion attack the speech signal of a source speaker is converted using some algorithm to match to the speech signal of a target speaker. The first ASVspoof2015 challenge [4] was mainly comprised of voice conversion and speech synthesis spoofing attacks. In the course of the challenge and subsequently, it became clear that the most effective countermeasures against voice conversion and speech synthesis spoofing attacks are low-level acoustic features designed to detect artifacts in synthetic or converted speech. By selecting a suitable feature [6, 7] it is possible achieve an equal error rate (EER) less than 1% on the evaluation set of ASVspoof2015 challenge data. Based on above finding it can be said that voice conversion and speech synthesis attacks are detectable too by reliable speaker verification systems. The last spoofing attack is replay or playback attack in which the voice of a target speaker is recorded using recording equipment and later played it to recognition system to spoof the system. Liveness detection (where the user is asked to utter a randomly selected sentence) is found to effective in this situation [1, 5].

Because of its simplicity and spreading of low cost recording devices and smart phones, at present, replay attacks pose a real threat to speaker verification systems. In order to draw interest among researchers to develop robust countermeasures for replay attack detection the second ASVspoof2017 challenge was conducted last year [9]. Before ASVspoof2017 challenge only a few research works were done on replay attack [8, 10, 11, 12]. In [13] and [23] replay attacks for text-dependent speaker verification were investigated. The focus of ASVspoof2017 challenge is also text-dependent speaker verification - replay attacks.

For the ASVspoof2017 challenge task various countermeasures using signal processing and deep learning techniques have been proposed and evaluated. In [14], [15] and [17] deep learning architectures in tandem with low-level acoustic features are used to detect replay attacks present in the ASVSpoof2017 challenge corpus. In [16], a fused system based on different types counter measures with various classifiers is presented for replay attack detection. Over-fitting problem caused by speech signal variability is investigated in [18]. In [19] and [20], a single frequency filtering technique and generalized teager energy operator-based instantaneous frequency features, respectively, are introduced for replay attack detection. Investigative studies are conducted in [21] and [22] on several low-level features and on different steps of the system, respectively. In [24], constant Q cepstral coefficients (CQCC) [7] feature-based feature selection technique is introduced to avoid high variance and over-fitting problems.

In the ASVspoof2017 challenge, feature normalization is found to play a key role to improve replay attack detection performance. Hence, in this work, we use DFT-based spectral

10.21437/Odyssey.2018-55

and product spectral features [6, 25] and propose to employ an intermediate domain, known as q-log domain [26], between the linear and log domain to normalize these features. By normalizing spoofing countermeasures in this way we were able improve replay attack detection performance. Final features are obtained by applying PCA instead of DCT to the log spectrum and product spectrum. The q-log function has already been used to extract robust features for speech recognition [27, 28] but not in speaker recognition and spoofing detection tasks.

Our main contribution in this work is two-fold: the first is the normalization of spoofing countermeasures in an intermediate domain using non-extensive statistics, and the second is the use of product spectrum for replay attack detection. Besides we show the importance high frequency features (frequency above 4 kHz) for relay spoofing detection task.

The remainder of this paper is organized as follows: We begin with a description of discrete Fourier transform-based power spectral and product spectral features that employ q-log domain feature normalization and a brief overview of q-log function. This is followed by a short description about baseline systems considered for this work and the backend used for classification. We then proceed to our experimental setups and results. We conclude with some concluding remarks.

2. FEATURES EXTRACTION

In this section, we describe discrete Fourier transform – based power spectral and product spectral features that utilize q-log domain feature normalization.

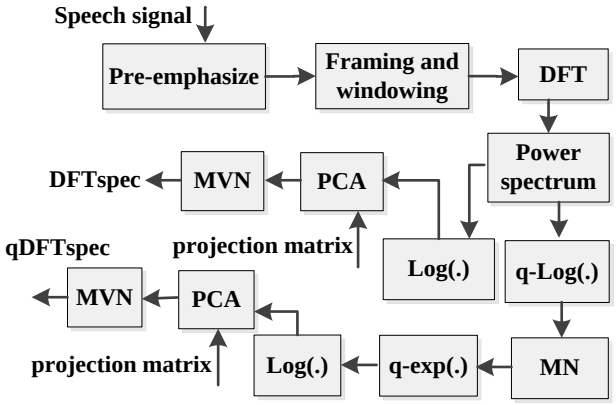

Fig. 1. Block diagram showing various steps to extract DFT-based log spectral (DFTspec) and q-log domain mean normalized log spectral (qDFTspec) features. The dimension of the features is reduced using principal component analysis (PCA). The PCA projection matrix is trained on the ASVspoof2017 training data. MN and MVN stands for mean normalization and mean and variance normalization, respectively.

If $X(m,k)$ is the Fourier transform of a framed and windowed time domain signal $x(m,n)$, $Y(m,k)$ is the Fourier transform of $y(m,n)=nx(m,n)$, where n is the sample index, m is frame index and k is the frequency bin index, then the power spectrum is $S(m,k)=|X(m,k)|^2$ and the product spectrum $P(m,k)$ is computed as [25, 29]:

$$P(m,k)=X_R(m,k)Y_R(m,k)+X_I(m,k)Y_I(m,k), \quad (1)$$

where subscripts R and I denote the real and imaginary parts, respectively.

Following the computation of power spectrum **DFTspec** and **qDFTspec** features are computed using the procedures mentioned in fig. 1. Similarly, based on product spectrum **Pspec** and **qPspec** features are calculated following the steps mentioned in fig. 2.

In the case **qDFTspec** and **qPspec** features an additional spectral mean normalization (MN) based on non-extensive statistics is performed in the *q*-log domain which acts as an intermediate domain between linear and log domains. By normalizing features in this way the nonlinear relation between speech and noise may be captured which is not well represented in log spectrum domain normalization. After performing feature normalization, the spectra are transformed back to linear domain and logarithm compression is applied.

Final features are obtained by using mean and variance (MVN) normalization technique followed by principal component analysis. The dimension of features after applying PCA is 90. No delta or double coefficients are added. The analysis window length is 20 ms with a frame shift of 10 ms.

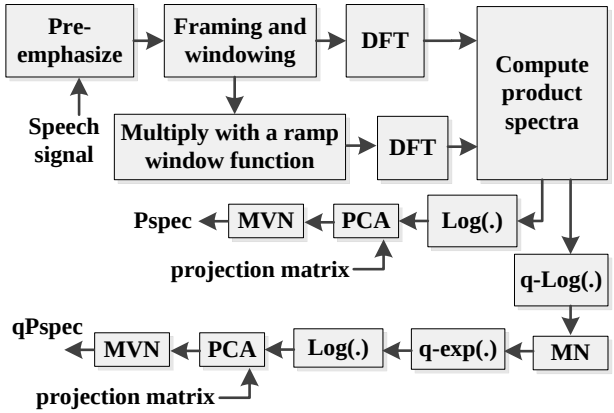

Fig. 2. Block diagram showing various steps of the Discrete Fourier transform (DFT)-based log product spectral features (Pspec) and q-log domain mean normalized log product spectral features (qPspec). Principal component analysis (PCA) is used to reduce feature dimension. The PCA projection matrix is trained on the ASVspoof2017 training data. MN and MVN stands for mean normalization and mean and variance normalization, respectively.

2.1. The *q*-log function

The *q*-log function, also known as generalized logarithm, is defined as [26-28]:

$$\log_q(x) = \begin{cases} \dfrac{x^{(1-q)}-1}{1-q} & \text{if } q \neq 1 \\ \log(x) & \text{if } q = 1 \end{cases}, \quad (2)$$

where q is a real number which should selected properly. In this work, we empirically set q = 0.94.

The q-exp (inverse of q-log) is defined as:

$$\exp_q(x) = \begin{cases} \left(1+(1-q)x\right)^{\left(\frac{1}{1-q}\right)} & \text{if } q \neq 1 \\ \exp(x) & \text{if } q = 1. \end{cases} \quad (3)$$

The q-log function has non-additivity property; therefore, it does not transform a multiplication into addition. For example,

$$\log_q(xy) = \log_q(x) + \log_q(y) + \ldots (1-q)\log_q(x)\log_q(y), \quad (4)$$

$$\log_q\left(\frac{x}{y}\right) = \frac{\log_q(x) - \log_q(y)}{1+(1-q)\log_q(y)}. \quad (5)$$

For more detail about q-log and q-exp functions and their properties please see [30].

2.2. Features normalization in q-log domain

So, in q-log domain the mean normalization of power spectrum is performed using following formula [30]:

$$\bar{S}_q(m,k) = \frac{\log_q\left(S(m,k)\right) - \frac{1}{N}\sum_{j=1}^{N}\log_q\left(S(m,k)\right)}{1+(1-q)\frac{1}{N}\sum_{j=1}^{N}\log_q\left(S(m,k)\right)}, \quad (6)$$

where $\bar{S}_q(m,k)$ is the mean normalized power spectrum in the q-log domain. To transform $\bar{S}_q(m,k)$ back to linear domain (3) is used.
Following the same procedure, the product spectrum is mean normalized in q-log domain and then transformed back to linear domain for further processing.

3. BASELINE SYSTEMS

The constant Q cepstral coefficients (CQCC) [7], linear predictive cepstral coefficients (LPCC), linear frequency cepstral coefficients (LFCC) features [6, 23, 29] and DFT-based log power spectrum (DFTspec) features are used as countermeasures for baseline systems. These features are also use a mean and variance normalization technique for feature normalization. The final feature dimension for CQCC, LFCC and LPCC front-ends is 60 including the delta and double coefficients. For DFTspec the feature dimension is 90. We use 20 ms analysis window length with a frame shift of 10 ms.

4. BACKEND

In this work, as backend we employ a standard Gaussian Mixture Model (GMM) classifier to distinguish playback signals from genuine speech signals. The 512-component GMM is used for training the genuine speech and spoof speech models. Let λ_h and λ_s represent the GMM models for genuine and spoof speech signals, respectively. Then, given the feature vector sequence O of a test speech signal, the genuine or spoof speech is decided based on the following log-likelihood ratio:

$$\ell(O) = \log\frac{\left(p(O|\lambda_s)\right)}{\log\left(p(O|\lambda_h)\right)}. \quad (7)$$

To train genuine and replay spoof models two training conditions are considered. In the first condition, models are trained only on the features extracted from the provided official training data. We denote this as condition 1. In the other condition the models are trained on the combined train plus dev features.

5. PERFORMANCE EVALUATION

In this section, we evaluate the performances of q-log normalized DFT spectral (qDFTspec) and product spectral (qPspec) features in terms of equal error rates (EERs). Replay spoofing detection experiments are conducted on the version 2 of ASVspoof2017 corpus and results are reported on the evaluation set. Development test set is used for tuning the parameters of the systems. Performances of qDFTspec and qPspec are compared with the baseline systems mentioned in section 3.

5.1. The ASVspoof2017 challenge corpus

In this work, the experiments are conducted on the version 2 of ASVspoof2017 challenge corpus [31] which is derived from the RedDots corpus [9, 23]. In version 2, zero-sequence artifacts in the beginning of the wave files and two empty files from training set have been removed.

The entire database is divided into three non-overlapping subsets, namely, train, dev and eval. More detail about this database can be found in [9, 23, 31]. Table 1 presents subsets of ASVspoof2017 corpus (version 2) in terms of number of speakers and utterances.

Table 1. Subsets of version 2 of the ASVspoof2017 challenge corpus in terms of number of speakers and number of utterances.

		# utterances		
Subset	# speakers	Genuine	Replay	Total
train	10	1507	1507	3014
dev	8	760	950	1710
eval	24	1298	12008	13306
Total	42	3565	14465	

5.2. Results and Discussion

According to challenge protocol the metric used for performance evaluation is equal error rate (EER). The results are reported on the evaluation set of ASVspoof2017 corpus (version 2).

In table 2 we present EERs obtained by all standalone replay spoofing attack detection systems developed for this work. The influence of feature normalization on replay attack detection is evident by observing the performance of CQCC with feature normalization (i.e., **CQCC**) and without feature normalization (i.e., **CQCC (w/o MVN)**).

The product spectral features (**Pspec**) which incorporates both amplitude and phase spectra, shows slightly better performance than DFT-based spectral (**DFTspec**). Comparing the performances of **DFTspec** versus **qDFTspec** and **Pspec** versus **qPspec** it is observed that *q*-log domain mean normalization is helpful for improving replay attack detection performance.

Table 2. Results of the standalone replay spoofing attack detection systems on the evaluation set of version 2 of ASVspoof2017 challenge data in terms of EER.

EER (%)		
Systems	**Training conditions**	
	Train	**Train + Dev**
DFTspec	12.12	11.5
qDFTspec	**11.43**	**11.19**
Pspec	12.08	11.38
qPspec	11.85	**11.23**
CQCC	22.82	16.58
CQCC (90)	23.85	17.94
CQCC (w/o MVN)	29.74	22.34
LFCC	15.76	13.77
LPCC	17.0	15.02

Though performance improvement is less with **qPspec** compared to **Pspec** the **qDFTspec** features provides a relative improvement of approximately 5% over **DFTspec** features. These performances indicate the benefit of feature normalization in *q*-log domain. All four spectral and product spectral features (with and without *q*-log spectral mean normalization) demonstrated better performance compared to the CQCC, LFCC and LPCC features.

Among all countermeasures the **qDFTspec** countermeasure showed the lowest EER. When the genuine and spoof models are trained only on the ASVspoof2017 training data, the **qDFTspec** system resulted in relative improvements of 5%, 49%, 27%, and 32% over the baseline DFTspec, CQCC, LFCC and LPCC systems, respectively. On the average **qDFTspec** provided a relative improvement of approximately 28% over all four baseline systems.

In order to show the influence of q-log domain spectral mean normalization and high frequency spectral regions i.e., spectral regions above 4 kHz, on replay attack detection task, we developed following two systems on the top of LFCC system:

- **qLFCC**: This is similar to qDFTspec but instead of PCA a discrete cosine transform is applied for decorrelation. After that first 20 coefficients are selected, including the 0^{th} cepstrum, as static features. Including the delta and double delta coefficients the final dimension of this features is 60.
- **qLFCC_LH:** Extraction of qLFCC_LH features is similar to qLFCC but unlike qLFCC 30-dimensional static features are formed by concatenating first 15 and last 15 coefficients. By doing so we build a system that incorporate both low and high frequency regions of the spectrum. After appending delta and double delta coefficients the final features dimension becomes 90.

Table 3. Influence of q-log domain spectral mean normalization and high frequency regions on replay spoofing attack detection task employing LFCC features. Results are reported in terms of EER on the evaluation set of version 2 of ASVspoof2017 challenge data.

EER (%)		
	When models are trained on training data	**When models are trained on (training + development) data**
LFCC	15.76	13.77
qLFCC	15.4	13.70
qLFCC_LH	14.37	13.15

Results presented in table 3 demonstrate the influence of q-log domain spectral mean normalization and high frequency spectral regions using LFCC features on the replay spoofing detection task. It is observed from this table that normalization of spectral mean in the q-log domain helped to reduce the error rate in both training conditions. Inclusion of high frequency regions yielded further reduction in EER in both training conditions. When models are trained only on provided training data, with qLFCC_LH features we achieved a relative improvement of 8.8% over LFCC features.

6. CONCLUSION

In this work, we investigated discrete Fourier transform (DFT) - based spectral and product spectral features for the detection of replay attacks on the second edition of ASVspoof challenge conducted in 2017. As feature normalization was found very useful for replay attack detection, for this task, we employed an additional spectral mean normalization based on non-extensive statistics in the *q*-log domain. The *q*-log function acts as an intermediate domain between linear and log domains. Performing feature normalization in this fashion may be able to capture the nonlinear relation between speech and noise signals, which is not well represented in log spectrum domain normalization based on extensive statistics. The DFT-based spectral and product spectral features, with and without a *q*-log spectral mean normalization, provided better performances than the baseline systems. Among all features the qDFTspec feature yielded the lowest EER.

7. ACKNOWLEDGEMENTS

This work has been made possible by investments from Ministry of Economic, Science and Innovation (MESI), Government of Quebec, Canada.

8. REFERENCES

1. Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi, Zhizheng Wu, Federico Alegre and Phillip De Leon, "Speaker recognition anti-spoofing," in the *Handbook of Biometric Anti-spoofing*, Springer, S. Marcel, S. Li and M. Nixon, Eds., 2014.
2. T. Kinnunen, Z. Wu, K. A. Lee, F. Sedlak, E. S. Chng, and H. Li, "Vulnerability of speaker verification systems against voice conversion spoofing attacks: The case of telephone speech," in International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4401-4404, 2012.
3. P. L. De Leon, M. Pucher, and J. Yamagishi, "Evaluation of the vulnerability of speaker verification to synthetic speech," in Proc. IEEE Speaker and Language Recognition Workshop (Odyssey), pp. 151-158, 2010.
4. Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilçi, M. Sahidullah, A. Sizov, "ASVspoof 2015: the First ASV Spoofing and Countermeasures Challenge," in proc. of INTERSPEECH, 2015. http://www.spoofingchallenge.org/is2015_asvspoof.pdf
5. Nanxin Chen, Yanmin Qian, Heinrich Dinkel, Bo Chen, Kai Yu, "Robust Deep Feature for Spoofing Detection - The SJTU System for ASVspoof 2015 Challenge", in proc. of Interspeech, 2015.
6. Jahangir Alam, Patrick Kenny, "Spoofing Detection Employing Infinite Impulse Response - Constant Q Transform-based Feature Representations," in proc. of EUSIPCO, Kos Island, Greece, 2017.
7. M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in *Speaker Odyssey Workshop*, Bilbao, Spain, 2016.
8. J. Villalba and E. Lleida, "Preventing replay attacks on speaker verification systems," in IEEE International Carnahan Conference on Security Technology, pp. 284–291, 2011.
9. T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi, and K. A. Lee, "The ASVspoof2017 challenge: Assessing the limits of replay spoofing attack detection," in proc. of Interspeech, Stockholm, Sweden, 2017.
10. F. Alegre, A. Janicki, and N. Evans, "Re-assessing the threat of replay spoofing attacks against automatic speaker verification," in proc. of International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–6, Sept 2014.
11. P. Korshunov, S. Marcel, H. Muckenhirn, A. R. Gonalves, A. G. S. Mello, R. P. V. Violato, F. O. Simoes, M. U. Neto, M. de Assis Angeloni, J. A. Stuchi, H. Dinkel, N. Chen, Y. Qian, D. Paul, G. Saha, and M. Sahidullah, "Overview of BTAS 2016 speaker anti-spoofing competition," in proc. of 8th International conference on Biometrics Theory, Applications and Systems, Sept 2016.
12. H. Muckenhirn, M. Magimai.-Doss, and S. Marcel, "Presentation attack detection using long-term spectral statistics for trustworthy speaker verification," in proc. of International Conference of the Biometrics Special Interest Group (BIOSIG), Sep. 2016.
13. Z. Wu, S. Gao, E. S. Cling, and H. Li, "A study on replay attack and anti-spoofing for text-dependent speaker verification," in proc. of Signal and Information Processing Association Annual Summit and Conference (APSIPA), Asia-Pacific, Dec 2014.
14. Parav Nagarsheth, Elie Khoury, Kailash Patil, Matt Garland, "Replay Attack Detection using DNN for Channel Discrimination," in proc. of Interspeech, Stockholm, Sweden, 2017.
15. Galina Lavrentyeva, Sergey Novoselov, Egor Malykh, Alexander Kozlov, Oleg Kudashev, Vadim Shchemelinin, "Audio Replay Attack Detection with Deep Learning Frameworks," in proc. of Interspeech, Stockholm, Sweden, 2017.
16. Zhe Ji, Zhi-Yi Li, Peng Li, Maobo An, Shengxiang Gao, Dan Wu, Faru Zhao, "Ensemble Learning for Countermeasure of Audio Replay Spoofing Attack in ASVspoof2017," in proc. of Interspeech, Stockholm, Sweden, 2017.
17. Zhuxin Chen, Zhifeng Xie, Weibin Zhang, Xiangmin Xu, "ResNet and Model Fusion for Automatic Spoofing Detection," in proc. of Interspeech, Stockholm, Sweden, 2017.
18. Lantian Li, Yixiang Chen, Dong Wang, Thomas Fang Zheng, "A Study on Replay Attack and Anti-Spoofing for Automatic Speaker Verification," in proc. of Interspeech, Stockholm, Sweden, 2017.
19. K.N.R.K. Raju Alluri, Sivanand Achanta, Sudarsana Reddy Kadiri, Suryakanth V. Gangashetty, Anil Kumar Vuppala, "SFF Anti-Spoofer: IIIT-H Submission for Automatic Speaker Verification Spoofing and Countermeasures Challenge 2017," in proc. of Interspeech, Stockholm, Sweden, 2017.
20. Hemant A. Patil, Madhu R. Kamble, Tanvina B. Patel, Meet H. Soni, "Novel Variable Length Teager Energy Separation Based Instantaneous Frequency Features for Replay Detection," in proc. of Interspeech, Stockholm, Sweden, 2017.
21. Roberto Font, Juan M. Espín, María José Cano, "Experimental Analysis of Features for Replay Attack Detection -Results on the ASVspoof 2017 Challenge," in proc. of Interspeech, Stockholm, Sweden, 2017.
22. Weicheng Cai, Danwei Cai, Wenbo Liu, Gang Li, Ming Li, "Countermeasures for Automatic Speaker Verification Replay Spoofing Attack: On Data Augmentation, Feature Representation, Classification and Fusion," in proc. of Interspeech, Stockholm, Sweden, 2017.
23. T. Kinnunen, M. Sahidullah, M. Falcone, L. Costantini, R. Gonzalez Hautam¨aki, D. Thomsen, A. Sarkar, Z.-H. Tan, H. Delgado, M. Todisco, N. Evans, V. Hautam¨aki, and K. Aik Lee, "RedDots replayed: A new replay spoofing attack corpus for text-dependent speaker verification research," in ICASSP, 2017.
24. Xianliang Wang, Yanhong Xiao, Xuan Zhu, "Feature Selection Based on CQCCs for Automatic Speaker Verification Spoofing," in proc. of Interspeech, Stockholm, Sweden, 2017.
25. D. Zhu and K. Paliwal, "Product of power spectrum and group delay function for speech recognition," in Proc. Int.

Conf. Acoust., Speech, Signal Process. , pp. 125–128, 2004.
26. C. Tsallis, “Possible generalization of boltzmann-gibbs statistics,” Journal of Statistical Physics, vol. 52, pp. 479–487, 1988.
27. H. F. Parade, “On noise robust feature for speech recognition based on power function family,” in proc. of ISPACS, pp. 386-390, Nov 2015.
28. H., F., Parade, K., Iwano, K., Shinoda, “Features normalization based on non-extensive statistics for speech recognition,” in Speech communication, vol. 55 (5), pp. 587-599, 2013.
29. Md. Jahangir Alam, Patrick Kenny, Gautam Bhattacharya and Themos Stafylakis, "Development of CRIM System for the Automatic Speaker Verification Spoofing and Countermeasures Challenge 2015," Proc. Interspeech, Dresden, Germany, Sept. 2015.
30. L. Nivanen, A. L. Mehaute, Q. A. Wang, “Generalized algebra within a non-extensive statistics,” Reports of Mathematical Physics, vol. 52 (3), pp. 437-444, 2003.
31. Kinnunen, Tomi; Sahidullah, Md; Delgado, Héctor; Todisco, Massimiliano; Evans, Nicholas; Yamagishi, Junichi; Lee, Kong Aik, “The 2nd Automatic Speaker Verification Spoofing and Countermeasures Challenge (ASVspoof 2017) Database, Version 2,” University of Edinburgh. The Centre for Speech Technology Research (CSTR). http://dx.doi.org/10.7488/ds/2301.