# TRANSMISSION LINE COCHLEAR MODEL BASED AM-FM FEATURES FOR REPLAY ATTACK DETECTION

*Tharshini Gunendradasan[1], Saad Irtza[1], Eliathamby Ambikairajah[1,2], Julien Epps[1,2]*

[1]School of Electrical Engineering and Telecommunications, UNSW, Australia
[2]ATP Research Laboratory, DATA61, CSIRO, Australia

**ABSTRACT**

This paper focuses on providing a countermeasure to replay attack which is the simplest and more accessible form of attack used to spoof automatic speaker verification systems. Specifically, it proposes the use of the transmission line cochlear model, which resembles the human cochlea more accurately than parallel filter bank models, in the front-end of replay detection systems. Here the basilar membrane is modeled as a cascade of digital filters with decreasing resonant frequencies. In this context, we propose two features – transmission line cochlea-amplitude modulation (TLC-AM) and frequency modulation (TLC-FM) – to extract the modulation features of the speech from the simulated membrane displacements. TLC-AM is analogous to the output of the inner hair cell bending movement, which accurately captures the amplitude modulation component of the speech. TLC-FM is extracted by deriving the in-phase and out of phase signals of basilar membrane displacement. Results show that individual TLC-AM and TLC-FM features perform better than the best parallel filter bank baseline system. Experiments suggest that higher frequency selectivity is beneficial for replay detection, especially for AM, and the proposed TLC model is better able to achieve this property than parallel filter bank models.

***Index Terms***— Transmission line cochlear model, amplitude modulation, frequency modulation, spoofing, replay attack, speaker verification

## 1. INTRODUCTION

Automatic speaker verification (ASV) is a mature technology that uses voice biometric to verify a person's identity [1]. As speech can be assessed remotely and deployment of speaker verification is simple and cost-effective, it has been adopted by several applications for secure verification e.g. telephone banking, physical access control. Although current ASV systems verify the identity with high accuracy and low equal error rate (EER), their vulnerability to spoofing attack has been shown to be significant [2], and dramatically affects the reliability and security of the system.

There are four main types of spoofing attacks currently under consideration: replay [3], speech-synthesis (SS) [4], voice conversion (VC) [5] and impersonation [6]. Replay attacks involve playing back the recorded speech of the genuine target speaker to spoof the system. Among all these attacks, replay poses the biggest threat due to the availability of high-quality recording devices and smartphones and the non-requirement of any advanced technical knowledge or effort [2]. This paper focuses on providing countermeasures for replay spoofing attack, to verify either the given speech is genuine or replayed.

Different countermeasures have been recently proposed for replay detection. Mel filter bank slope and linear filter bank slope features [7] have been proposed to capture low and high frequency spectral information separately. In [8], static and dynamic characteristics of the modulation spectrum were fused with short time magnitude features to improve the system performance. Linear prediction based features were proposed in [9], [10]. Spectral centroid based amplitude and frequency modulation features were proposed in [11]. Apart from these front-end features, different neural network architectures e.g. convolutional neural networks and Siamese network have been proposed which are effective in discriminating replay attacks [12], [13], [14].

Replayed utterances contain both additive and convolutional distortions introduced by recording and playback devices [15]. As replay attacks involve multiple recording and playback, they will be affected by noise [16]. The amplitude-based features of the signal can capture these distortions. It has been suggested that the changes in spectral envelope due to the channel characteristic of intermediate devices can be captured by phase-based features [17]. As instantaneous information can capture the dynamic and time evolution of the speech features, instantaneous amplitude and phase-based features can effectively capture these distortions. The contribution of these two features in discriminating genuine and replayed speech has been exploited in the following past work. In [18] auditory filter banks were learned using ConvRBM and AM and FM features were extracted using the conventional energy separation algorithm. VESA-IACC and VESA-IFCC features were proposed in [19], where instantaneous amplitude (IA) and instantaneous frequency (IF) were estimated using the VESA algorithm from the Gabor filtered subband signal. Moreover, the importance of these two features for detecting SS and VC was also analyzed in [20].

The above-mentioned methods use parallel filter banks to obtain the subband signals for instantaneous amplitude and phase extraction. The cochlear model in the human/mammalian auditory system, which inherently has extraordinary frequency sensitivity and selectivity, can be more accurately approximated by a transmission line model [21]. In the transmission line model, the cochlea is represented as a cascade of digital filters, which helps to achieve sharper roll off even with smaller order filters, allowing high frequency selectivity [21]. The importance of choice of filters with different shape in discriminating genuine and replayed speech was reported in [22]. Motivated by these, we hypothesize that analyzing the AM and FM components of the signal with high frequency sensitivity and resolution would be effective in capturing differences between genuine and spoofed speech. Thus, we proposed transmission line cochlear model to extract the AM and FM components of the speech. We refer to the proposed features as transmission line cochlea AM and FM (TLC-AM, TLC-FM).

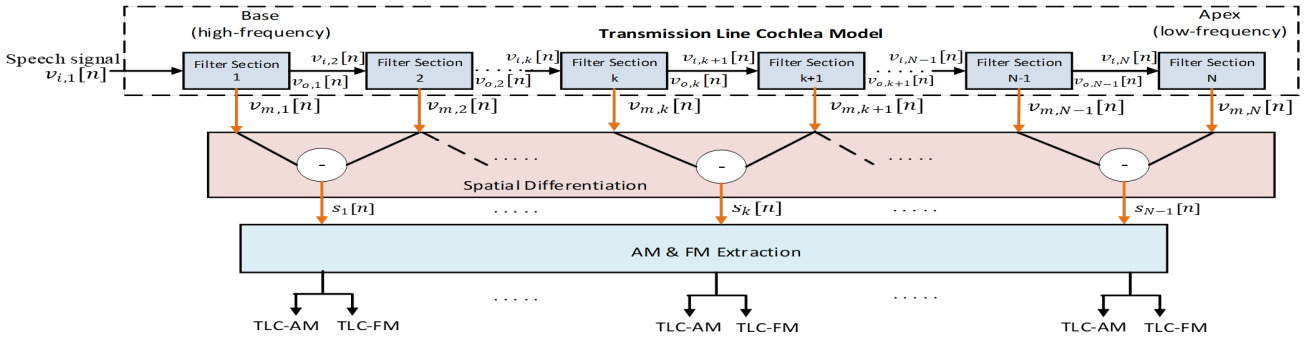


**Figure 1.** Block diagram of transmission line model of cochlea as a cascade of digital filters, showing the process of extracting TLC-AM and TLC-FM features from the basilar membrane displacement. $v_{m,k}[n]$ and $s_k[n]$ denote the membrane displacement and spatially differentiated displacement, respectively.

## 2. PROPOSED TLC-AM AND TLC-FM FEATURES

### 2.1. Transmission Line Model of the Cochlea

We used the transmission line cochlea model as previously proposed in [23]. In this model, the wave propagation in the cochlea is modelled as a cascade of filter sections, as shown in Figure 1. The input stimuli travel down along the basilar membrane from base to apex, that is, from high frequency to low frequency, and the membrane displaces according to the frequency content of the stimuli. The input to each digital filter section is the pressure, and that is converted into displacement of the basilar membrane, and the output pressure is transmitted to the following filter section. This can be described in two transfer functions.

The first transfer function is the pressure transfer function: for the $k^{\text{th}}$ filter section it relates the input pressure ($v_{i,k}[n]$) and the output pressure ($v_{o,k}[n]$), where $n$ denotes time/sample. For a single filter section, this can be represented as the cascade of a lowpass filter, resonant filter and notch filter, respectively [24]:

$$H_k(z) = \frac{V_{o,k}(z)}{V_{i,k}(z)} = K\,\frac{1-a_0}{1-a_0 z^{-1}}\,\frac{1-b_1+b_2}{1-b_1 z^{-1}+b_2 z^{-2}}\,\frac{1-a_1 z^{-1}+a_2 z^{-2}}{1-a_1+a_2} \quad (1)$$

where $K$ is a gain factor, and $a_0, a_1, a_2, b_1$ and $b_2$ are the digital filter coefficients. The lowpass filter at the $k^{\text{th}}$ filter section passes the frequencies below the digital resonant frequency of section $k$ ($\theta_{ck}$) to pass low-frequency pressure towards the apex to vibrate, the resonant filter resonates at $\theta_{ck}$ and the notch filter removes the frequencies above $\theta_{ck}$ to stop the pressure passing to the next filter section. The overall response of the filter section $k$ is the cascade of all the filters preceding it.

$$G_k(z) = \prod_{i=1}^{k} H_i(z) \quad (2)$$

The next transfer function is the displacement transfer function, relating input pressure ($v_{k,i}[n]$) to output displacement ($v_{k,m}[n]$):

$$\frac{V_{k,m}(z)}{V_{k,i}(z)} = K\,\frac{1-a_0}{1-a_0 z^{-1}}\,\frac{1-b_1+b_2}{1-b_1 z^{-1}+b_2 z^{-2}} \quad (3)$$

In (1), it can be observed that the pressure transfer function contains the displacement transfer function. Thus, the cochlear model can be designed by simply cascading the filters.

In the cochlear model, the membrane displacement is spatially differentiated to model the fluid coupling between the adjacent sections of the basilar membrane, thus providing additional sharpening mechanisms in the basilar membrane.

$$s_k[n] = v_{m,k+1}[n] - v_{m,k}[n] \quad (4)$$

where $s_k[n]$ and $v_{m,k}[n]$ are the spatially differentiated displacement and membrane displacements for section $k$.

The frequency response of the $k^{\text{th}}$ spatially differentiated displacement is analogous to a bandpass filter. Figure 2 compares the bandpass filter response of the proposed transmission line cochlear model with the gamma tone filter. It can be observed that a sharp roll off is achieved with the proposed model compared with the gammatone filter, ensuring the high frequency selectivity of the proposed cascaded model.

### 2.2. TLC-AM Feature Extraction

The process of extracting AM components from the spatially differentiated displacement $s_k[n]$ is briefly illustrated in Figure 1. In the human auditory system the inner hair cell, which performs the mechanical (membrane displacement) to neural transduction process in the cochlea, effectively captures the AM component of the signal. We propose this AM component as a feature for spoofing detection and refer to it as the transmission line cochlea AM (TLC-AM).

The inner hair cell model we implemented is a capacitive model composed of a half-wave rectifier followed by a lowpass filter [25]. The process of extracting the TLC-AM feature, which approximates the action of the inner hair cell, is shown in Figure 3. The spatially differentiated signal $s_k[n]$ is passed to the half wave rectifier, and the rectified signal $p_k[n]$ is then lowpass filtered to obtain the TLC-AM feature $f_k[n]$:

$$\frac{F(z)}{P(z)} = \frac{1-c_0}{1-c_0 z^{-1}} \quad (5)$$

where $c_0$ is the digital filter coefficient.

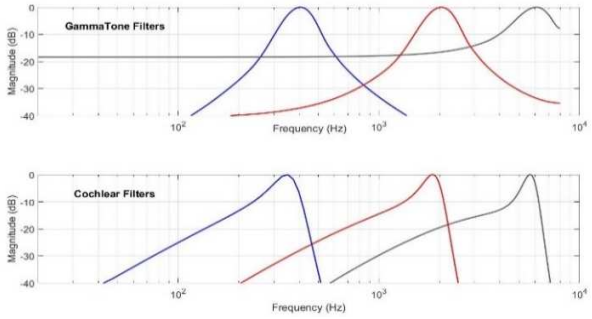


**Figure 2.** Frequency response of Gammatone filter and the proposed transmission line cochlear model.

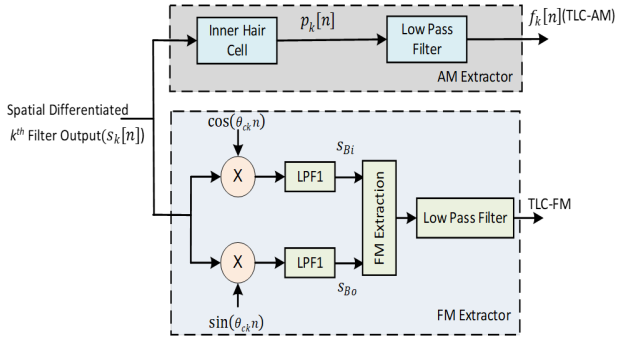


**Figure 3.** Process of extracting TLC-AM and TLC-FM from the $k^{\text{th}}$ spatially differentiated signal ($\boldsymbol{s_k[n]}$).

### 2.3. TLC-FM Feature Extraction

The FM extraction from the spatially differentiated displacement $s_k[n]$ is shown in Figure 1. The FM component of the speech signal is extracted similarly to the algorithm used in [26], [27]. The advantage of using this algorithm over the conventional Hilbert transform based FM extraction is that in the Hilbert transform method, the FM components tend to vary very rapidly and vary within a broad range, resulting in difficulties in providing clear physical meaning [27]. This problem is addressed in the proposed algorithm by extracting band limited and slowly varying FM [27].

The process of FM extraction from the spatially differentiated signal $s_k[n]$ is illustrated in Figure 3. In the first step, if $s_k[n]$ is represented as $S_k[n] = a[n]\cos(\theta_{ck}n + \varphi[n])$, the cosine and sine modulated signals are estimated by multiplying by the orthogonal sine and cosine signals with digital frequency $\theta_{ck}$. These modulated signals are then low pass filtered with a cutoff frequency chosen to be the same as the bandwidth of the $k^{th}$ cochlear filter section, to decompose the modulated signals into in-phase ($s_{Bi}$) and out-of-phase ($s_{Bo}$) signals of $s_k[n]$:

$$s_{Bi} = \frac{1}{2}a[n]cos(\varphi[n]) \quad (6)$$

$$s_{Bo} = -\frac{1}{2}a[n]sin(\varphi[n]) \quad (7)$$

The FM components are then extracted as follows [27]:

$$FM = \frac{s_{Bo}\,\Delta\, s_{Bi} - s_{Bi}\,\Delta s_{Bo}}{2\pi(\,{s_{Bi}}^2 + {s_{Bo}}^2)} \quad (8)$$

Then the extracted FM component is lowpass filtered to remove high frequency distortions that appear during instantaneous FM extraction [28]. We refer to this FM feature as Transmission line cochlear FM (TLC-FM).

## 3. EXPERIMENT SETTINGS

### 3.1 Database

In this paper, our proposed method was evaluated on ASV spoof 2017 version 2 replay corpus [28], , which was released after the first version (V1) [29] to correct some of the anomalies present in that, which may impact the assessment of replay detection. Both of the databases were derived from the RedDots corpus used for ASV [30].

Most of the results reported recently for replay attack are on V1 database, thus for the comparison purpose results were reported both on V1 and V2 database for the proposed method.

### 3.2 Feature Extraction and Model Training

We conducted extensive experiments to choose the design parameters and filter coefficients of the cochlear model. We tested this model with Mel and linear scales for replay detection and a linear scale was found to give better results. The speech signals were pre-emphasized to emphasize the high frequency region as the high frequency regions are more discriminative in replay detection [31], [11]. For the feature extraction both TLC-AM and TLC-FM were averaged over frame sizes of 2ms and 20ms respectively, with an overlap of 50%. The reason for the small frame size for TLC-AM compared with the typical frame size used in replay detection is that the cut-off frequency of the lowpass filter in the hair cell model influences the selection of the frame size. The log compression was performed only for TLC-AM. Then discrete cosine transform (DCT) and mean-variance normalization (MVN) were performed on both TLC-AM and TLC-FM.

In our experiments it was observed that for TLC-FM features, including delta features alongside the static features was helpful, while for TLC-AM features all delta, delta-delta coefficients were useful in discriminating genuine and replay attack. Thus, in all our experiments we employed this setting.

GMM was used as the back-end classifier and spoofed and genuine classes were modelled using 512 mixture components from the training and development data. For the test utterances the classification scores were calculated as the log likelihood ratio between genuine and spoofed classes. We used the equal error rate (EER) as the metric to evaluate the performance of the system, which is the primary metric used in all other spoofing detection systems. For the fusion of two systems, Focal multi-class toolkit [32] is used to perform liner score-level fusion.

### 3.3 Baseline features

We selected our baseline systems from existing time domain instantaneous AM and FM feature extraction methods that use parallel filter banks. This allowed us to compare the effectiveness of the proposed cascaded filter bank model relative to parallel filter bank approaches. For the baseline system, we also chose the methods that use linearly scaled filters. Our first baseline employed the VESA-IACC and VESA-IFCC [19] features, which use linearly scaled parallel Gabor and Butterworth filter banks for AM and FM extraction, respectively. The second baseline was the AM-ConvRBM-CC and FM-ConvRBM-CC features [18], which extract AM and FM components from the auditory filters learned from a Convolutional Restricted Boltzmann Machine (ConvRBM), where the learned filter banks were nearly linear scaled.

## 4. RESULTS AND DISCUSSION

### 4.1 The Effect of Higher Frequency Selectivity of TLC Model for AM and FM features

To analyse the impact of higher frequency selectivity achieved by TLC model in AM and FM features for replay detection, TLC-AM and TLC-FM are compared with gammatone parallel filterbank model that extract both features using the similar method proposed in this paper. The EER obtained for varying number of filters/ filter sections are shown in Figure 4. For both AM and FM features the EER is lower for TLC model than gammatone filters, especially AM showed significant reduction. Moreover, the minimum EER occurs for a large number of filters, implying that the discriminative information for replay attack stays within the small frequency bins.

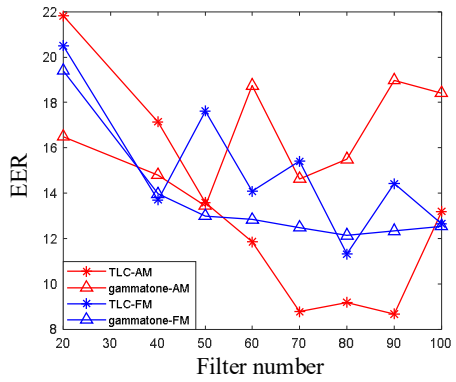


**Figure 4**: Variation of equal error rate with filter number for AM and FM features extracted from TLC model and gammatone filtered signal.

Compared to parallel filter bank models that have low/moderate frequency selectivity, the high frequency selectivity offered by TLC model effectively contributes to capture information within small frequency bins. Further, considering the relative improvement obtained by AM compared to FM for TLC model suggests that the higher selectivity is more beneficial for AM feature than FM for replay detection. As the minimum EER is obtained with 90 and 80 filters sections for AM and FM, respectively, we chose those values for all subsequent experiments.

### 4.2 Comparison of TLC Model with other Parallel Filter Bank AM and FM Feature Extraction Methods

Table 1 compares the results of TLC-AM and TLC-FM features with other parallel filter bank-based AM and FM feature extraction methods that use different filter banks such as Gabor, gammatone and Butterworth. For AM extraction, among parallel filter bank models Gabor and ConvRBM perform better than Butterworth filter. This observation can be related to the properties of the filters, noting that the Butterworth filter has flat passband frequency response whereas the others have optimum localization property in the frequency domain, again implying the importance of frequency selectivity for AM extraction for replay detection. TLC-AM outperforms other AM feature extraction methods and achieved an improvement of 28.7% relative to the best reported feature VESA-IACC that uses Gabor filter. This suggests that the proposed TLC-AM feature that gives higher frequency selectivity together with more accurate AM extraction using proposed Hair cell model contributes to capture the discriminative information present in the small frequency bins, this in turn helps to discriminate genuine and replayed speech. For FM extraction, according to the results reported for parallel filter bank models, the type of filterbanks doesn't make significant differences as in AM. Proposed TLC-FM feature showed an improvement of 14.3% relative to VESA-IFCC feature.

### 4.3 Results with Score-Level Fusion

As the amplitude and phase features capture two distinct types of information about speech sub band signals, the possible complementary nature of those two features was explored by fusing them at the score level. Table 2 compares the fusion results of the proposed TLC-AM + TLC-FM features with the baseline systems that use parallel filter banks for AM and FM extraction. The database organizers of ASV spoof version 2 provided CQCC as the baseline system [28], which was also included in this comparison. Our proposed best result was obtained after fusing TLC-AM and TLC-FM at score level, an improvement of 14% relative to the system based on TLC-AM features alone suggesting the complementary nature of TLC-AM and TLC-FM.

**Table 1:** Equal error rate results on the evaluation set for systems with AM and FM features extracted using the proposed TLC model and other parallel filter bank models using Butterworth (BW), Gabor (GA) and ConvRBM (CV) filters.

| | | Feature set | EER |
|---|---|---|---|
| AM | Parallel | HT-IACC [22]) (BW), (V1) | 19.27 |
| | | HT-IACC [22] (GA), (V1) | 12.12 |
| | | ESA-IACC [22] (BW), (V1) | 21.43 |
| | | ESA-IACC [22] (GA), (V1) | 12.00 |
| | | VESA-IACC [19] (GA), (V1) | 11.94 |
| | | AM-ConvRBM-CC [18] (CV), (V1) | 12.76 |
| | **TLC** | **TLC-AM (V1)** | **8.51** |
| | | **TLC-AM (V2)** | **8.68** |
| FM | Parallel | HT-IFCC [22] (BW), (V1) | 39.40 |
| | | HT-IFCC [22] (GA), (V1) | 14.62 |
| | | ESA-IFCC [22] (BW), (V1) | 28.69 |
| | | ESA-IFCC [22] (GA), (V1) | 12.79 |
| | | VESA-IFCC [19] (BW), (V1) | 11.79 |
| | | FM- ConvRBM-CC [18] (CV), (V1) | 14.96 |
| | **TLC** | **TLC-FM (V1)** | **10.11** |
| | **TLC** | **TLC-FM (V2)** | **11.30** |

**Table 2:** Replay detection equal error rate on the evaluation set for score-level fusion of previous and proposed systems in comparison with baseline system using CQCC features.

| Features Set | EER |
|---|---|
| VESA-IACC+ VESA-IFCC [19], (V1) | 7.11 |
| AM-ConvRBM-CC + FM-ConvRBM-CC [18], (V1) | 8.89 |
| **TLC-AM+ TLC-FM (V1)** | **7.32** |
| **TLC-AM+ TLC-FM (V2)** | **7.59** |
| HT-IACC+ HT-IFCC [22], (V1) | 10.03 |
| ESA-IACC + ESA-IFCC [22], (V1) | 9.64 |
| CQCC [28], (V2) | 12.24 |

## 5. CONCLUSION

In this paper we proposed the application of the transmission line cochlea model, which models the basilar membrane as a cascade of digital filters, applying high-frequency selectivity to the replay attack detection problem. Two features, TLC-AM and TLC-FM were proposed to extract the amplitude and frequency modulation components of the speech, respectively and were compared with other parallel filter bank-based AM and FM extraction methods. As seen from the experimental results, the fusion of the proposed TLC-AM and TLC-FM features match the best previous AM-FM approaches, however individually both the TLC-AM and TLC-FM features show significantly improved performance over other individual AM and FM features which can be attributed to higher frequency selectivity of the TLC model, which enables discriminative information extraction even in small frequency bins for AM and FM features and helps to distinguish replayed speech from genuine speech. Especially, AM features obtained with higher frequency selectivity TLC model was more beneficial for replay detection.

## 6. REFERENCES


[1] Hansen JH, Hasan T.: 'Speaker recognition by machines and humans: A tutorial review,' *IEEE Signal processing magazine*, vol. 14, pp. 74-99, 2015.

[2] Wu, Z., Evans, N., Kinnunen, T., Yamagishi, J., Alegre, F., and Li, H.: 'Spoofing and countermeasures for speaker verification: a survey,' A survey. *speech communication*, *66*, pp.130-153, 2015.

[3] Gałka, J., Grzywacz, M., and Samborski, R.: 'Playback attack detection for text-dependent speaker verification over telephone channels,' Speech Communication, 67, pp. 143-153, 2015.

[4] Zen, H., Tokuda, K., and Black, A.W.: 'Statistical parametric speech synthesis,' Speech Communication, *51*(11), pp.1039-1064, 2009.

[5] Correia, M., Abad, A., and Trancoso, I.: 'Anti-spoofing: Speaker verification vs. voice conversion,' stituto Superior Técnico Master's Thesis, 2014

[6] Lau, Y.W., Wagner, M., and Tran, D.: 'Vulnerability of speaker verification to voice mimicking,' *2004. Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 145-148, 2004.

[7] M S, S., Murthy, H.: 'Decision-level Feature Switching as a Paradigm for Replay Attack Detection,' *Proc. Interspeech*, pp.686-690, 2018. DOI: 10.21437/Interspeech.2018-1494

[8] Suthokumar, G., Sethu, V., Wijenayake, C., and Ambikairajah, E.: 'Modulation Dynamic Features for the Detection of Replay Attacks,' *Proc. Interspeech,* pp. 691-695, 2018.

[9] Wickramasinghe, B., Irtza, S., Ambikairajah, E., and Epps, J.: 'Frequency Domain Linear Prediction Features for Replay Spoofing Attack Detection,' *Proc. Interspeech*, pp. 661-665, 2018.

[10] Jelil, S., Kalita, S., Prasanna, S.R.M., Sinha, R.: 'Exploration of Compressed ILPR Features for Replay Attack Detection,' *Proc. Interspeech*, pp 631-635, 2018. DOI: 10.21437/Interspeech.2018-1297.

[11] Gunendradasan, T., Wickramasinghe, B., Le, N.P., Ambikairajah, E., and Epps, J.: 'Detection of Replay-Spoofing Attacks Using Frequency Modulation Features,' *Proc. Interspeech*, pp. 636-640, 2018.

[12] Tom, F., Jain, M., and Dey, P.: 'End-To-End Audio Replay Attack Detection Using Deep Convolutional Networks with Attention,' *Proc. Interspeech*, pp. 681-685, 2018.

[13] Sriskandaraja, K., Sethu, V., and Ambikairajah, E.: 'Deep Siamese Architecture Based Replay Detection for Secure Voice Biometric,' *Proc. Interspeech*, pp. 671-675, 2018.

[14] Lavrentyeva, G., Novoselov, S., Malykh, E., Kozlov, A., Kudashev, O., Shchemelinin, V.: 'Audio Replay Attack Detection with Deep Learning Frameworks,'*Proc.Interspeech*, pp 82-86, 2017.

[15] Rafi, B.S.M., Murty, K.S.R., and Nayak, S.: 'A new approach for robust replay spoof detection in ASV systems,' IEEE *GlobalSIP,* pp. 51-55, 2017.

[16] Raju Alluri, K., and Gangashetty, A.K.V.: 'SFF Anti-Spoofer: IIIT-H Submission for Automatic Speaker Verification Spoofing and Countermeasures Challenge 2017,' *InProc. Interspeech,* pp. 107-111, 2017.

[17] Patil, H.A., Kamble, M.R., Patel, T.B., and Soni, M.: 'Novel Variable Length Teager Energy Separation Based Instantaneous Frequency Features for Replay Detection,' *Proc. Interspeech*, pp. 12-16, 2017.

[18] Sailor, H., Kamble, M., and Patil, H.: 'Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection,' *Proc. Interspeech,* pp. 666-670, 2018.

[19] Kamble, M., and Patil, H.: 'Novel Variable Length Energy Separation Algorithm Using Instantaneous Amplitude Features for Replay Detection,' *Proc. Interspeech*, pp. 646-650, 2018.

[20] Patel, T.B., and Patil, H.A.: 'Combining evidences 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*. 2015.

[21] Kates, J.M.: 'Accurate tuning curves in a cochlear model,' *IEEE Transactions on Speech and Audio Processing*, vol. 4, pp. 453-62, 1993.

[22] Kamble, M., Tak, H., and Patil, H.: 'Effectiveness of Speech Demodulation-Based Features for Replay Detection,' *Proc. Interspeech*, pp. 641-645, 2018.

[23] Ambikairajah, E., Black, N.D., and Linggard, R.: 'Digital filter simulation of the basilar membrane,' *Computer Speech and Language*, vol. 3, pp. 105-118, 1989.

[24] Ambikairajah, E., and Kilmartin, L.: 'An Adaptive Cochlear Model for Speech Recognition,' *In Second European Conference on Speech Communication and Technology*,1991.

[25] Shamma, S.A., and Morrish, K.A.: 'Synchrony suppression in complex stimulus responses of a biophysical model of the cochlea,' *The Journal of the Acoustical Society of America*, vol. 5, pp. 1486-1498, 1987.

[26] Flanagan, J.L., and Golden, R.: 'Phase vocoder,' *Bell System Technical Journal*, vol. 9, pp. 1493-1509, 1966.

[27] Nie, K., Stickney, G., and Zeng, F.-G.: 'Encoding frequency modulation to improve cochlear implant performance in noise,' *IEEE transactions on biomedical engineering*, vol.1, pp. 64-73, 2005.

[28] Delgado, H., Todisco, M., Sahidullah, M., Evans, N., Kinnunen, T., Lee, K.A., and Yamagishi, J.: 'ASVspoof 2017 Version 2.0: meta-data analysis and baseline enhancements,' In *Proc. Odyssey The Speaker and Language Recognition Workshop*, pp. 296-303, *2018*.

[29] T. Kinnunen *et al*.: 'ASVspoof 2017: automatic speaker verification spoofing and countermeasures challenge evaluation plan', *Training,* vol. 10, no. 1508, p. 1508, 2017.

[30] Lee, K.A., Larcher, A., Wang, G., Kenny, P., Brümmer, N., Leeuwen, D.v., Aronowitz, H., Kockmann, M., Vaquero, C., and Ma, B.: 'The RedDots data collection for speaker recognition', In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[31] Witkowski, M., Kacprzak, S., Zelasko, P., Kowalczyk, K., and Gałka, J.: 'Audio Replay Attack Detection Using High-Frequency Features', *Proc. Interspeech*, pp. 27-31, 2017.

[32] Brümmer N.: 'Focal multi-class: Toolkit for evaluation, fusion and calibration of multi-class recognition scores tutorial and user manual,' http://sites.google.com/site/nikobrummer/focalmulticlass, 2007.