Novel Energy Separation Based Instantaneous Frequency Features for Spoof Speech Detection

Madhu R. Kamble and Hemant A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar-382007, Gujarat, India
{madhu_kamble, hemant_patil}@daiict.ac.in

***Abstract*—Speech Synthesis (SS) and Voice Conversion (VC) presents a genuine risk of attacks for Automatic Speaker Verification (ASV) technology. In this paper, we evaluate front-end anti-spoofing technique to protect ASV system for SS and VC attack using a standard benchmarking database. In particular, we propose a novel feature set, namely, Energy Separation Algorithm-based Instantaneous Frequency Cosine Coefficients (ESA-IFCC) to detect the genuine and impostor speech. The experiments are carried out on ASV Spoof 2015 Challenge database. On the development set, the score-level fusion of proposed ESA-IFCC feature set with Mel Frequency Cepstral Coefficients (MFCC) gave an EER of 3.45 %, which reduced significantly from MFCC (6.98 %) and ESA-IFCC (5.43 %) with 13-D static features. The EER decreases further to 2.01 % and 1.89 % for Δ and $\Delta\Delta$ features derived from proposed ESA-IFCC features, respectively. The overall average error rate for known and unknown attacks in evaluation set was 6.79 % for ESA-IFCC and was significantly better than the MFCC (9.15 %) features.**

***Index Terms*—Teager Energy Operator, Energy Separation Algorithm, Instantaneous Frequency Cosine Coefficients, SSD, GMM, EER.**

I. Introduction

Automatic Speaker Verification (ASV) or voice biometrics is the task to authentic the claimed person's individuality from his or her voice with the help of machines [1]. However, practical ASV systems are vulnerable to biometric attacks known as *spoofing*. The major forms of attacks known today includes, voice conversion, speech synthesis, identical twins, replay, and impersonation, which are known to degrade performance of ASV systems [1]. The general countermeasure approach is one of the solutions to focus on feature extraction and statistical pattern recognition techniques. In particular, feature extraction forms key task for spoof speech detection (SSD). The aim is to distinguish between genuine and impostor speech by capturing the key discriminative features between two speech signals. This might propose that front-end side should be focused more for countermeasures of spoofing than on the sophisticated classifiers. This view is supported by the results of the recent ASV Spoof 2015 Challenge, during INTERSPEECH 2015 [1]-[2].

Most of the ASV systems today use Fourier transform (FT) magnitude-based features from the speech signal. In general, universally used features for speech signal are Mel Frequency Cepstral Coefficients (MFCC) [3] and Linear Prediction Cepstral Coefficients (LPCC) [4] (i.e., implicit LP spectrum) that captures only magnitude information. The access to extract features from phase spectrum were not as suitable as magnitude spectrum counterpart of a signal [5] and [6]. However, the phase characteristics of speech are also found to be significant as the magnitude for speaker characterization [7], [8]. Studies have been reported to use Fourier transform phase-based features, such as, Modified Group Delay (MGD) features [9] and [10], temporal modulation [11] to detect genuine *vs.* impostor speech. Another way to decompose a signal, is by using Hilbert transform, it is the product of a slowly-changing envelope and a rapidly-changing fine time structure. This study indicates that the envelope carry significant information for speech reception, whereas for pitch perception and sound localization, the fine structure is important. When these two features clashes, location of a sound is determined by the fine structure, however, the words are analyzed by the envelope [12].

In this paper, we propose Energy Separation Algorithm-based Instantaneous Frequency Cosine Coefficients (ESA-IFCC) feature set. Energy Separation Algorithm (ESA) algorithm utilizes nonlinear Teager-Kaiser Energy Operator (TEO), that estimates energy of a monocomponent signal as the product of its squared amplitude and frequency [13]. If the frequency of a signal changes immediately then, the TEO algorithm effectively tracks the instantaneous frequency (IF) information. If the signal has a lot of wide range of frequency components (such as in speech signal), then the signal has to be bandpass subband filtered (as TEO works effectively on monocomponent signals) and then ESA is applied at the output of each subband signals. Temporal average is enumerated to obtain L-dimensional IF coefficients (IFCs) for each frame. Discrete Cosine Transform (DCT) is applied on IF to obtain stabilized cepstral features, i.e., proposed ESA-IFCC feature set [14]. Authors proposed to emphasize Fourier Transform (FT) phase information via IF computation using ESA, in ESA-IFCC. The performance of ESA-IFCC features is compared with MFCC for SSD task on ASV Spoof 2015 Challenge database.

II. Energy Separation Algorithm (ESA)

According to the Teager [15], the vortex-flow interactions are the true source of sound production, which are nonlinear

ISBN 978-0-9928626-7-1

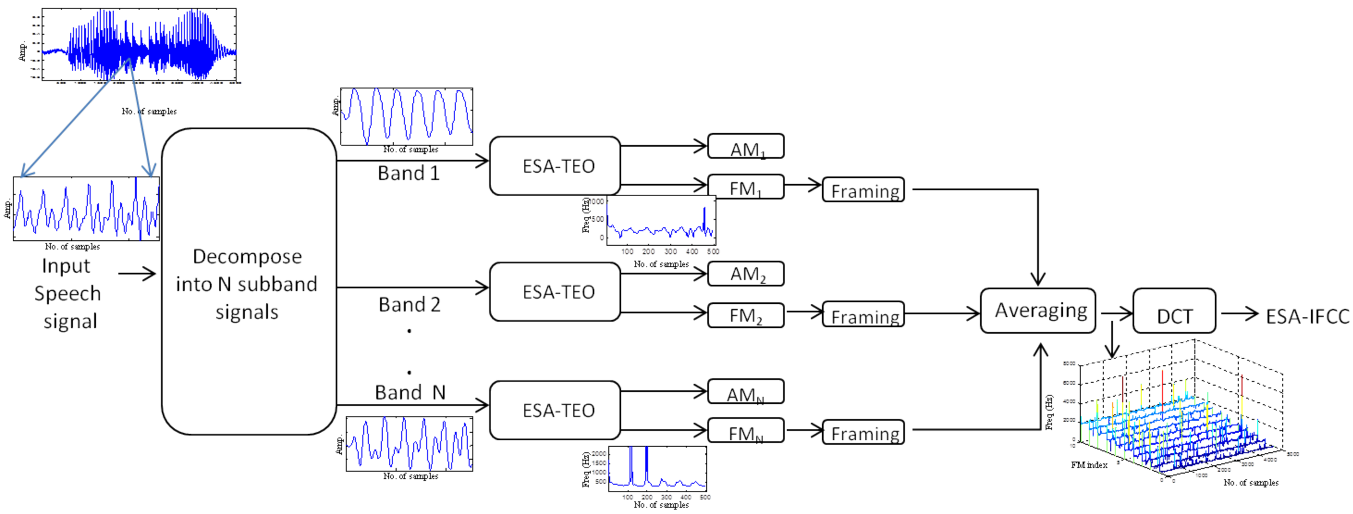

Fig. 1. Schematic diagram for extraction of proposed ESA-IFCC feature set. The 3-D plot before DCT block is corresponding to entire utterance.

in nature. Teager derived an algorithm, that uses a nonlinear energy tracking operator called as TEO (in discrete-time) for speech signal analysis, which was tough task to model the time-changing vortex-flow. The concept was further extended on continuous-time domain by Kaiser [13].

A. Teager-Kaiser Energy Operator

Newton's second law of motion for an oscillator with mass m and spring constant k describing following differential equation

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0, \tag{1}$$

and its solution consists of a sinusoidal signal $x(t) = Acos(\Omega t + \phi)$, where A is the amplitude of the oscillation, Ω is frequency of oscillations (rad/sec) and ϕ is the initial phase (rad). The sum of the potential and kinetic energy gives the system's total energy E, i.e.,

$$E = \frac{1}{2}kx^2 + \frac{1}{2}m\dot{x}^2 \Rightarrow E = \frac{1}{2}m\Omega^2 A^2, \tag{2}$$

where $\Omega = d\phi(t)/dt$. Taking this analysis under consideration, Teager and then Kaiser [13], proposed the TEO for discrete-time signal $x[n] = Acos(\omega n + \phi)$, i.e.,

$$E_n = \Psi_d\{x(n)\} = x^2(n) - x(n-1)x(n+1) \approx A^2\omega^2, \tag{3}$$

where E_n gives the running estimate of signal's energy and in $\Psi_d\{.\}$ d is for discrete-time case. A simple algorithm for nonlinear energy tracking operator is referred to as TEO for analysis of signal with the observation and belief that hearing is the process of detecting energy [13].

To estimate the individual contribution of amplitude $a[n]$ and frequency $\omega[n]$ of signal, Maragos et. al. [16], [17] developed an *energy separation algorithm (ESA)* using nonlinear energy operator that tracks the instantaneous energy of the AM-FM signal to separate the signal into its amplitude and frequency components. Both amplitude and frequency are function of energy in a speech signal [18]. However, ESA is carried out on single speech resonance, as speech itself is a combination of several resonances, and hence, we require to separate resonances with bandpass subband filtering. The IF $\omega[n]$ and AE $a[n]$ at any time instant of the AM-FM modulated signal $x[n]$ is given by [19]:

$$a[n] \approx \frac{2\Psi_d\{x[n]\}}{\sqrt{\Psi_d\{x[n+1]) - x[n-1]\}}}, \tag{4}$$

$$\omega_i[n] \approx arcsin\sqrt{\frac{\Psi_d\{x[n+1] - x[n-1]\}}{4\Psi_d\{x[n]\}}}. \tag{5}$$

Eq. (4) and Eq. (5) are w.r.t. symmetric approximation of derivative, i.e., $y[n] = \frac{x(n+1)-x(n-1)}{2}$. The frequency estimation part assumes that $0 < \omega_i[n] < \frac{\pi}{2}$ because the computer implementation of $arcsin(u)$ function assumes that $|u| < \frac{\pi}{2}$. Thus, this discrete ESA can be used to estimate IF $< 1/4$ the sampling frequency of signal [16]. The IF of a signal possibly explained as the frequency of the sinusoid that fits the given signal *locally*. The IF is modeled as the superposition of slow and fast-changing factor. The average formant frequency values are modeled by slow-changing factor and frequency variations around the formant frequency are modeled by fast-changing factor.

III. Proposed ESA-IFCC Features

Fig. 1 shows the block diagram of proposed ESA-IFCC feature set. Here, the input speech signal is first split into N frequency bands or subband signals. The ESA is then applied onto each N bandpass (subband) filtered signals to obtain corresponding AEs and IFs. Furthermore, we have discarded the AE and taken only IF. IFs are computed for each of the narrowband components. The IF are disjointed into short frames of 20 ms duration, with a shift of 10 ms, and obtained L-dimensional IF coefficients (IFCs) for each frame

ISBN 978-0-9928626-7-1 © EURASIP 2017

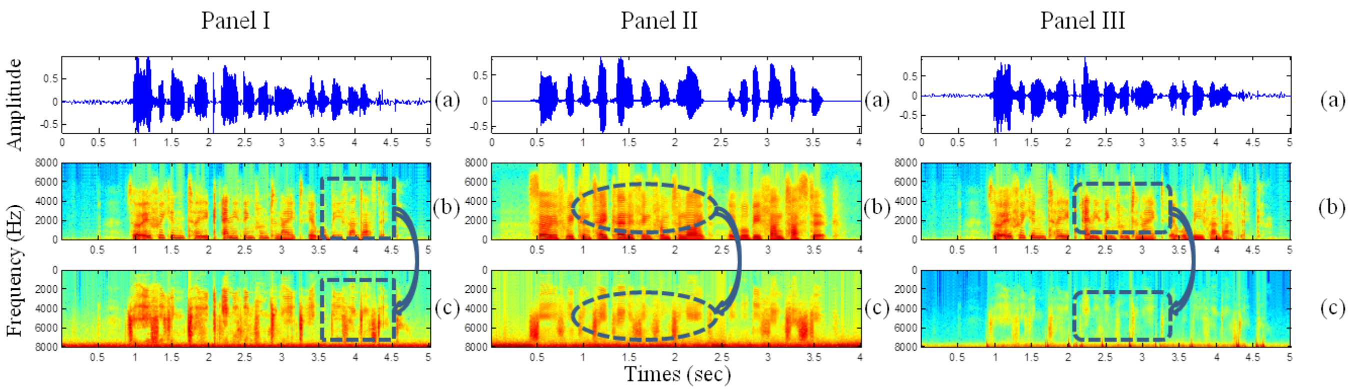

Fig. 2. Spectrographic analysis (a) speech signal (b) corresponding spectrogram (c) spectral energy density of 40 subband filtered signals (i.e., N=40 in Fig. 1). Panel I for natural speech, Panel II for SS speech, and Panel III for VC speech.

with temporal average computation. The redundancy among IFCs is exploited to obtain a low-dimensional representation by employing DCT (that has energy compaction property) and thus, retaining first few DCT coefficients. The low-dimensional features obtained by applying DCT on IF coefficients is referred to as ESA-based Instantaneous Frequency Cosine Coefficients (ESA-IFCC). IFCC along with their Δ and $\Delta\Delta$ features were appended to get higher-dimensional cosine features denoted as ESA-IFCC.

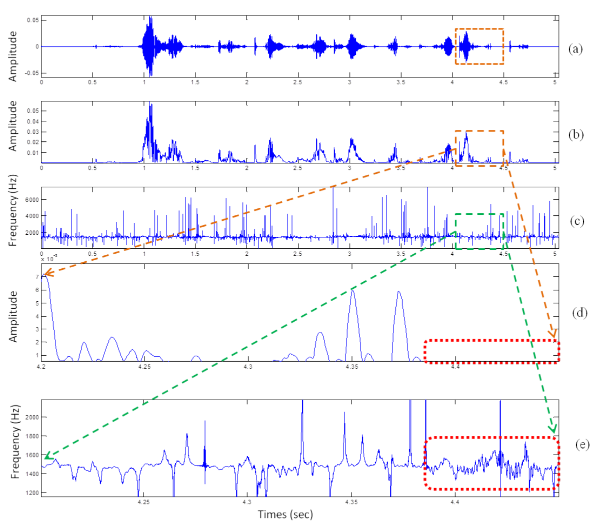

Fig. 3. AM-FM decomposition (a) speech signal, (b) AE, (c) IF of filtered narrowband signal at $f_c = 1500$ Hz (i.e., for N=8 in Fig. 1) and Fig. 3(d)-(e) amplitude envelope and instantaneous frequency, respectively, of speech segment shown with dotted box in Fig. 3(a).

Fig. 3(a) shows the plot of a genuine speech utterance and Fig. 3(b) and Fig. 3(c) shows the respective AE and IF of a narrowband filtered signal around 1500 Hz for a speech signal shown in Fig. 3(a). The output is possibly decomposed into its corresponding AE and IF from the filterbank of every narrowband component as shown in Fig. 1. The IF in Fig. 3(c) is centered around 1500 Hz and shows spurious fluctuations on both side that makes it difficult to analyze and interpret the vocal tract system characteristics [14]. There could be two reasons for IF that has spurious fluctuations in a speech signal and they are:

- When the amplitude approaches to zero, then the large fluctuation are been observed for IF. For the region 4.4 to 4.5 s Fig. 3(c), i.e., unvoiced region have more changes because of narrowband component has low energy in that region as shown in Fig. 3(d).
- On the other hand, the region from 4.2 to 4.25 s and 4.3 to 4.4 s as in Fig. 3(c), which is voiced regions have the changes in IF connected to impulse-like nature during speech production [14].

This is because the IF computed from the speech signal contains the information from both vocal tract system and excitation source. As a result, IF shows impulse-like discontinuities at the instants of glottal closure [20]. As a result, at the instants of Glottal Closure Instants (GCI), discontinuity of impulses are observed in IF. The impulse response of vocal tract system during production of a speech signal originated GCI successively to yield a speech signal. The phase discontinuity occur due to superposition of impulse response that gives us the large amplitude peaks in IF at locations of GCI.

A. *Butterworth filterbank*

The Butterworth filter provides a maximally flat response (i.e., the first $2n - 1$ derivatives for the power function w.r.t. frequency are zero and hence, has no ripples) in the passband. In this usually, Butterworth filterbank is used with filters placed according to linear frequency scale. Fig. 2 shows the spectrographic analysis of natural (panel I), SS (panel II) and VC (panel III) speech signals. Fig. 2(a) shows the time-domain speech signal and its corresponding spectrogram is shown in Fig. 2(b), whereas the spectrogram obtained after 40 subband Butterworth filtered signals is shown in Fig. 2(c). It can be observed that the higher frequency regions corresponding to the higher spectral amplitude (designating the vocal tract resonances, also referred to as formants) are more emphasized

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE I
RESULTS IN TERMS OF EER (%) ON DEVELOPMENT DATASET SCORE-LEVEL FUSION AS PER EQ. 7

Feature 1		# EER (%) for varying α												Feature 2
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	Frequency Range	
MFCC	static	6.98	6.72	6.36	6.00	5.69	5.40	**5.24**	5.30	5.54	6.63	8.16	100-3000 Hz 13-D	ESA-IFCC
	Δ	6.75	6.25	5.84	5.36	4.83	4.42	4.13	3.88	**3.85**	4.27	5.29		
	$\Delta\Delta$	6.14	5.71	5.29	4.81	4.31	3.98	3.74	**3.70**	3.85	4.56	5.79		
MFCC	static	6.98	6.28	5.67	5.05	4.48	4.17	3.91	**3.80**	4.06	4.78	6.38	100-7800 Hz 40-D	ESA-IFCC
	Δ	6.75	4.05	2.76	2.27	**2.18**	2.34	2.56	3.28	4.17	5.67	7.47		
	$\Delta\Delta$	6.14	2.98	2.17	**1.98**	2.12	2.42	3.00	3.63	4.62	5.75	7.18		
MFCC	static	6.98	6.57	6.13	5.67	5.18	4.66	4.16	3.75	**3.45**	3.62	5.43	100-7800 Hz 13-D	ESA-IFCC
	Δ	6.75	5.58	4.50	3.72	3.14	2.71	2.30	**2.01**	2.02	2.71	6.22		
	$\Delta\Delta$	6.14	5.10	4.08	3.31	2.82	2.41	2.02	**1.89**	2.00	2.76	6.59		

in natural speech, while it is less emphasized for SS speech and for VC speech, i.e., the formants in spoofed speeches are not at all visible. Hence, it could be the speaker-specific information that can make the natural, SS and VC distinguishable.

The filter must be as wide as possible to include the desired formant modulations. However, narrow enough to exclude the interference of neighboring formants. The center frequencies of the bandpass filters are linearly-spaced and used to extract the component AM-FM signals of the speech segment and then determine the modulations around these center frequencies. Authors have chosen linearly-spaced filterbank for Butterworth filter as opposed to other frequency scale as Mel scale, Equivalent Rectangular Bandwidth (ERB) scale (this is in line with the recent finding reported in [14]).

IV. EXPERIMENTAL RESULTS

In this paper, we use ASV Spoof 2015 Challenge database that was created for the ASV spoofing and countermeasures challenge, and it comprises of genuine and impostor speech data [1]. Brief details of database are given in [1], [21]. The ESA-IFCC features were extracted using 40 channel linearly-scaled 3^{rd} order Butterworth filterbank for frequency range of 100-7800 Hz. For each narrowband component, TEO-based ESA is applied for computing AM-FM components. IF were computed for each of the narrowband components. Furthermore, these IFs were averaged over short-time windows of 20 ms duration, shifted by 10 ms, to obtain 13-D ESA-IFC features. The ESA-IFCC were obtained by applying DCT on IFCs and retaining the first 13 coefficients in the transformed-domain. The 13-D ESA-IFCCs, together with their dynamic and acceleration features were used to build the SSD system. We have used Gaussian Mixture Model (GMM) with 128 mixtures for modeling the classes corresponding to genuine and impostor speech utterances. Final scores are represented in terms of Log-Likelihood Ratio (LLR). The decision of the test speech being genuine or impostor is based on the LLR, i.e.,

$$LLR = log(LLK\_Model1) - log(LLK\_Model2), \quad (6)$$

where LLK_Model1 and LLK_Model2 are the likelihood scores from the GMM for the genuine and impostor trials, respectively. To explore the possible complementary information captured by MFCC and ESA-IFCC features, we use their score-level fusion, i.e.,

$$LLK_{combine} = (1-\alpha)LLK_{feature1} + \alpha LLK_{feature2}, \quad (7)$$

where α equal to weight of fusion.

A. Results on Development Dataset

Results for MFCC and proposed ESA-IFCC feature set are shown in Table I. From the results, proposed feature set captures speaker-specific information embedded in natural speech (as SS and VC speech does not exactly match the human speech) and hence, there exists differences between natural *vs.* spoofed classes. Table I shows that the ESA-IFCC features produce much lower % Equal Error Rate (EER) than the MFCC alone, ESA-IFCC features that are capable to distinguish genuine *vs.* impostor speech (i.e., for SS and VC, the features are comparable different than those for human speech).
It was found that linearly-scaled equi-spaced filters are more suitable for IF estimation than the ERB-scaled varying bandwidth filters. In the case of gammatone filterbank, the bandwidth increases at high frequencies, making the estimation of IF less reliable. However, authors have used Butterworth filter that has nonlinear phase that can be approximated as linear over smaller frequency regions. For given 16 kHz sampling frequency, we have available bandwidth of 7800 Hz that is divided into 40 equi-spaced frequency regions of width $(f_H - f_L)/40$ Hz. The phase response around each $(f_H - f_L)/40$ Hz width is mostly found to be linear (as observed in authors recent study reported in [22]).

Furthermore, the score-level fusion of these features was done as per Eq. (7) and is shown in Table I. It was observed that for almost equal weighted fusion of MFCC and ESA-IFCC score, the % EER of MFCC (6.98) and ESA-IFCC (5.43) reduces to 3.45 % for static features similar pattern was observed for Δ and $\Delta\Delta$ for higher frequency range of 100-7800 Hz. The contribution of a particular system is decided by the weight of fusion (i.e., α). From Table I, it is observed that most of the system contribution was done with ESA-IFCC features as the parameter α most of time was biased towards ESA-IFCC. Therefore, it can be said that the ESA-IFCC features have more contribution in decreasing the % ERR and hence, could classify in a more better way for SSD.

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE II
RESULTS IN % EER ON EVALUATION DATASET FOR EACH SPOOFING ATTACK. BOTH KNOWN AND UNKNOWN ATTACKS. +:SCORE-LEVEL FUSION

Features	Known Attacks						Unknown Attacks						All Avg.
	S1	S2	S3	S4	S5	Avg.	S6	S7	S8	S9	S10	Avg.	
A:MFCC	2.34	9.57	**0.00**	**0.00**	9.01	4.18	7.73	4.42	0.3	5.17	52.99	14.12	9.15
B:ESA-IFCC	2.68	4.87	**0.00**	**0.00**	12.87	4.08	10.9	2.4	3.57	3.33	**37.37**	**9.514**	**6.79**
A+B	**0.78**	**3.39**	**0.00**	**0.00**	**5.45**	**1.92**	**4.19**	**1.22**	**0.11**	**1.80**	54.73	**12.41**	7.16

From Table I, the proposed features capture the *complementary* information that was not observed from MFCC alone. The % EER, is very less for score-level fusion of the MFCC and ESA-IFCC.

B. Results on Evaluation Dataset

Table II shows the results for evaluation dataset with known and unknown spoofing attacks. It was observed that SS attacks (S3, S4) were easily detected for known attacks while S10 (MARY TTS) in unknown attacks was most difficult task to detect. These results show that performance degrades significantly with unknown attacks. The average performance for the unknown attacks is dominated by the performance for S10 then the performance of known attacks. The overall average error rate for known and unknown was 6.79 % for ESA-IFCC and was significantly better than the MFCC (9.15 %) features. The score-level fusion (when performed with the fusion factor $\alpha = 0.8$) gave the overall average EER of 7.16 % due to dominance of S10 unknown attack. However, with score-level fusion of MFCC and ESA-IFCC, other attacks from S1 to S9 (known and unknown attacks) were detected reasonably well.

V. SUMMARY AND CONCLUSIONS

This study presented the effectiveness of ESA-IFCC feature set in capturing speaker-specific information. The computation of IF is mainly affected by the parameters of the filterbank, namely, number of channels, shape of subband filters, etc. The proposed ESA-IFCC features shows the improvement on combining with MFCC features to detect genuine *vs.* impostor speech. The use of IF to capture perceptual information proves to be very effective. It was observed that on the standard dataset provided for the challenge, the score-level fusion of the MFCC and ESA-IFCC features gave quite low % EER on development dataset than the MFCC alone. The authors would like to explore use of ESA-IFCC features for robustness in presence of additive or channel noise and its relative effects of various types of spoofed speech.

REFERENCES

[1] 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.

[2] T. B. Patel and H. A. Patil, "Cochlear filter and instantaneous frequency based features for spoofed speech detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 618–631, 2017.

[3] 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, 1980.

[4] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[5] L. D. Alsteris and K. K. Paliwal, "Short-time phase spectrum in speech processing: A review and some experimental results," *Digital Signal Processing*, vol. 17, no. 3, pp. 578–616, 2007.

[6] J. W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1215–1247, 1993.

[7] L. D. Alsteris and K. K. Paliwal, "Further intelligibility results from human listening tests using the short-time phase spectrum," *Speech Communication*, vol. 48, no. 6, pp. 727–736, 2006.

[8] K. K. Paliwal and L. D. Alsteris, "On the usefulness of STFT phase spectrum in human listening tests," *Speech Communication*, vol. 45, no. 2, pp. 153–170, 2005.

[9] B. Yegnanarayana and H. A. Murthy, "Significance of group delay functions in spectrum estimation," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2281–2289, 1992.

[10] Z. Wu, C. E. Siong, and H. Li, "Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition,," in *INTERSPEECH, Portland, Oregon, USA*, 2012, pp. 1700–1703.

[11] Z. Wu, X. Xiao, E. S. Chng, and H. Li, "Synthetic speech detection using temporal modulation feature," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, BC, Canada, 2013, pp. 7234–7238.

[12] Z. M. Smith, B. Delgutte, and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, no. 6876, pp. 87–90, 2002.

[13] J. F. Kaiser, "On a simple algorithm to calculate the energy of a signal," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Albuquerque, New Mexico, USA, 1990, pp. 381–384.

[14] K. Vijayan, P. R. Reddy, and K. S. R. Murty, "Significance of analytic phase of speech signals in speaker verification," *Speech Communication*, vol. 81, pp. 54–71, 2016.

[15] H. Teager, "Some observations on oral air flow during phonation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 5, pp. 599–601, 1980.

[16] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Transactions on Signal Processing*, vol. 41, no. 10, pp. 3024–3051, 1993.

[17] Maragos, Petros and Kaiser, James F and Quatieri, Thomas F, "On separating amplitude from frequency modulations using energy operators," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, San Francisco, California, USA, 1992, pp. 1–4.

[18] A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth tracking using multiband energy demodulation," *The Journal of the Acoustical Society of America (JASA)*, vol. 99, no. 6, pp. 3795–3806, 1996.

[19] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Pearson Education India, 2006.

[20] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[21] Z. Wu, T. Kinnunen, N. W. D. Evans, J. Yamagishi, C. Hanilçi, M. Sahidullah, and A. Sizov, "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge,," in *INTERSPEECH*, Dresden, Germany, 2015, pp. 2037–2041.

[22] P. B. Bachhav, H. A. Patil, and T. B. Patel, "A novel filtering based approach for epoch extraction," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Brisbane, Australia: IEEE, 2015, pp. 4784–4788.