Subband Teager Energy Representations for Infant Cry Analysis and Classification

Ankur T. Patil, Aastha Kachhi, Hemant A. Patil
Speech Research Lab
Dhirubhai Ambani Institute of Information and Communication Technology
Gandhinagar, Gujarat, India
Email: {ankur_patil, aastha_kachhi, hemant_patil}@daiict.ac.in

***Abstract*—Due to quasi-periodic sampling of vocal tract system, spectrum formed by high pitch-source harmonics results in extremely poor spectral resolution for commonly used spectral features, such as Mel Frequency Cepstral Coefficients (MFCC). Therefore, classifying normal *vs.* pathological infant cry became technologically challenging signal processing problem. In that context, this study investigates the effectiveness of the spectral representations over cepstral representations for infant cry analysis. Furthermore, we show the spectrographic analysis for various spectral representations, such as Short-Time Fourier Transform (STFT), Mel Filterbank (MelFB) coefficients, Linear Filterbank (LinFB) coefficients, and proposed subband Teager Energies (subband-TE). Experiments are performed on standard Baby Chilanto dataset with Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) as classifiers. It is observed that spectral representations performs slightly better over the cepstral representations. Moreover, subband-TE representations with GMM classifier achieves 99.47% classification accuracy for normal *vs.* pathology cry and outperforms other architectures. This is due to the capability of TEO to accurately estimate the energies, especially in low frequency regions which consists of fundamental frequency (F_0) and its harmonics.**

***Index Terms*—Infant cry analysis and classification, pathological cry, Teager Energy Operator.**

I. Introduction

Crying is an infant's only mode of communication [1]. Millions of infants die within a few months of birth due to diseases, malnutrition, and vaccine-preventable diseases [2]. For this purpose, fingerprint and cry-based identification systems for neonates are being developed for infants [2], [3]. The most prevalent causes of infant's death are perinatal asphyxia, asthma, and Sudden Infant Death Syndrome (SIDS) [4], [5]. An arterial blood sample is necessary for numerous measures, including oxygen saturation, PH, and electrolyte, in order to clinically identify these disorders. Hence, detecting pathology takes longer time in developing nations and is costly, which results in death of infant. Asphyxia is one of those diseases that can be diagnosed by visual symptoms, like pale and bluish limbs. Nevertheless, by then, the newborn would have suffered significant neurological damage [5], [6]. Similarly, for the deaf infant, the acoustical features depend on the type of hearing loss, rehabilitation period, as well as the age at which the pathology was detected [7]. As a result, the demand for developing a cry diagnostic assistive tool to aid doctors in recognizing early indicators of such illnesses is growing. In the analysis and classification of infant cries, physiologists, neurologists, paediatricians, engineers, linguists, and psychologists are all involved. This paper presents a signal processing-based technique for classifying normal *vs.* pathological infant cries.

The initial work in this area started in early 1960s [8], where spectrogram was used. Various spectrographic cry modes, such as vibration, dysphonation, inhalation, hyperphonation, etc. were analysed in the healthy baby cries [9]. However, this study was extended for pathological cries in [10], where different cry modes were also found to be correlated in pathological cries. In this study, due to quasi-periodic sampling of vocal tract system, narrowband spectrogram barely shows the formant structures and hence, F_0 and its harmonics serve as the key features in this application as opposed to the other applications in speech signal processing.

Apart from the prosodic features, such as pitch F_0 [11], the state-of-the-art cepstral features, such as Mel Frequency Cepstral Coefficients (MFCC) with Gaussian Mixture Model (GMM) are also recently used for healthy *vs.* pathological infant cry classification [12], [13]. However, this paper investigates the effect of spectral feature for infant cry classification. Furthermore, the capability of the Teager Energy Operator (TEO) to accurately estimate the energies of the signal is exploited for the assigned task [14]. Teager Energy Cepstral Coefficients (TECC)-based on TEO are used originally for automatic speech recognition applications [15], [16]. Many studies reveal that the feature representation of the speech signal developed using TEO outperforms for SSD task [17], [18].

For this application, we proposed the subband Teager Energies (Subband-TE) for infant cry analysis. Subband-TE is equivalent to the spectral representation, as opposed to the cepstral. The reasoning behind the suitability of the spectral representations over cepstral for infant cry analysis is discussed in Section II. Furthermore, Section III presents the details of the proposed subband-TE and TECC feature sets. The experimental setup and results are discussed in Section IV and Section V, respectively. Finally, Section VI presents the summary and future scope of this work.

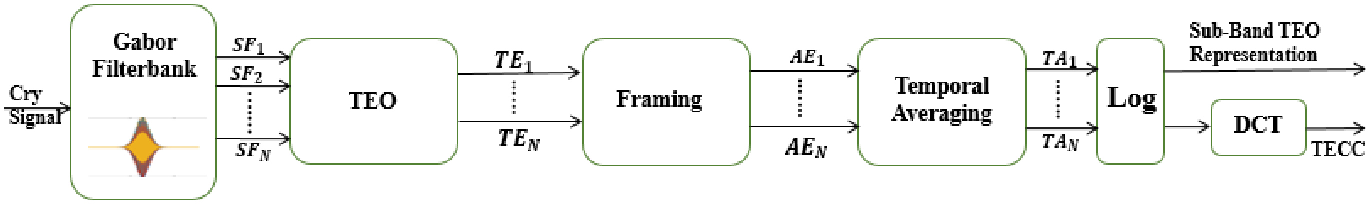

Fig. 1: Functional block diagram of the proposed Subband TEO representation and TECC feature set. (SF: Subband filtered signal, TE: Teager Energies, AE: Averaged Energies over frames, TA: Temporal Averaging). After [15], [19].

II. Spectral *vs.* Cepstral Analysis for Infants Cry

In speech signal processing applications, cepstral features are generally preferred over spectral features. Using linear filter theory, the speech signal $x(n)$ is modelled as the convolution of glottal airflow ($s(n)$) with vocal tract system ($h(n)$), i.e., $x(n) = s(n) * h(n)$, where '$*$' represents the convolution operation. In evaluation of the logarithmic cepstrum, convolutionally-combined vector space, $x(n) = s(n) * h(n)$, is mapped to the additively combined vector space, $\hat{x}(n) = \hat{s}(n) + \hat{h}(n)$, such that the contribution of glottal airflow $s(n)$, and impulse response of vocal tract system, $h(n)$ can be distinctly observed [20], [21]. This transformation takes place in such a way that the duration of the pulse-train, $\hat{s}(n)$ remains the same as that of $s(n)$, however, $\hat{h}(n)$ should get compressed (in quefrency-domain) than the $h(n)$ [20]. Here, $\hat{x}(n)$, $\hat{s}(n)$, and $\hat{h}(n)$ are referred to as logarithmic cepstrum of their corresponding time-domain signals, $x(n)$, $s(n)$, and $h(n)$, respectively. The logarithmic cepstrum is estimated as the inverse Fourier transform of the logarithm of the Fourier transform of the given signal $x(n)$, i.e.,

$$\hat{x}(n) = \mathcal{F}^{-1}(log(\mathcal{F}(x(n)))), \quad (1)$$

where $\mathcal{F}(\cdot)$ represents the Fourier transform. Because of the transformation given in eq. (1), convolutional vector space is transformed to additive vector space [22].

The source signal $s(n)$ is represented by the fundamental frequency (F_0), which varies w.r.t. age of the individual. In infants, the mass of the vocal folds is much lesser than that of an adult male. Hence, the frequency of the vibrations (and hence, F_0 and its harmonics) of the vocal folds in infants is much larger than that of adult male. The F_0 in infants and adult male is 250-400 Hz and 60-100 Hz, respectively. Due to relative differences in mass (and related inertia) of vocal tract system and vocal fold, frequency of oscillations of vocal folds is much greater than that of vocal tract system and hence, in quefrency domain, $\hat{s}(n)$ and $\hat{h}(n)$ are well separable in adult males, whereas for infants, it is difficult to separate. In various speech signal processing applications, we need to extract the vocal tract system information to extract feature set. Hence, cepstral representations are significant in those applications. In this work, the F_0 and its harmonics are being the cues for the classification of normal *vs.* pathology cry signals, spectral features are supposed to produce the desirable feature representation. This theoretical hypotheses is validated using the experiments, where spectral representations are shown to give better performance.

III. Subband Teager Energy Representations

According to conventional signal processing literature, energy of the speech signal $x(t)$ is estimated as $L^2(R)$-norm of the signal, i.e., square operation over the entire signal under analysis [23]. This approach of estimating the energy is based on linear filtering theory, which could model the linear components of the speech signal. However, speech production mechanism consists of the non-linearities and hence, the speech signal energies could not be estimated accurately using linear filter theory [24]. To alleviate this issue, TEO was introduced in [14]. It is a non-linear differential operator which could capture the non-linear aspect of the speech production mechanism and also the properties of airflow pattern in the vocal tract system [21], [23]. Energy of the given speech signal $x(t)$ having amplitude, A and monotone frequency, ω_m can be estimated using TEO as:

$$\Psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) = A^2\omega_m^2, \quad (2)$$

where $\dot{x}$ and $\ddot{x}$ represents the first and second-order derivative of the signal $x(t)$. The expression of the TEO in eq. (2) is derived from the solution of simple harmonic motion (SHM), which possess single frequency component. By approximating the derivative operation in continuous-time with backward difference in discrete-time, we obtain the TEO for discrete-time signal $x(n)$ having amplitude A and monotone angular frequency Ω_m as follows [14]:

$$\Psi[x(n)] = x^2(n) - x(n-1)x(n+1) \approx A^2\Omega_m^2. \quad (3)$$

TEO is derived to estimate the energy for monotone signal. However, speech signal consists of the frequency range varying from baseband to Nyquist frequencies. Hence, to obtain the monotone approximation of the signal, the speech signal is allowed to pass through the filterbank, which consists of several subband filters with appropriate center frequency and bandwidth. The subband filtered signals are narrowband signals which are supposed to approximate the monotone signals and hence, TEO can be applied on these subband filtered signals. In this work, Gabor filterbank with linearly-spaced subbandd filters, is utilized for subband filtering. TEO is applied on each subband filtered signal to accurately estimate the energy. Furthermore, these narrowband energies are segmented into the frames of 20 ms duration with overlapping of 10

ms. Then, the temporal average for each frame is estimated to produce N-dimensional (D) *subband Teager energy representations (subband-TE)*. Discrete Cosine Transform (DCT) is performed on *subband Teager energy representations* to obtain the TECC. The functional block diagram representation of the proposed subband-TE and TECC feature set is shown in Fig. 1. In this study, we analyzed the relative performance of the subband-TE *vs.* TECC feature set.

IV. Experimental Setup

A. Dataset Used

In this study, we utilized the standard Baby Chilanto database. It was designed using recordings made by doctors and is the property of Mexico's NIAOE-CONACYT [25]. Each cry signal was segmented into one second duration and then grouped into 5 groups. For binary classification of healthy *vs.* pathological cries, two groups were constructed, where healthy cry signals include normal, pain, and hunger cries totalling in 1049 cry samples. Pathological cry signals include asphyxia and deaf cries totalling in 1219 cry samples. The statistics of the dataset is shown in the Table I. As the number cry signals are less in number, experiments are performed using *10*-fold cross-validation.

TABLE I: Statistics of the Baby Chilanto dataset. After [25].

Class	Category	# Samples
Healthy	Normal	507
	Hunger	350
	Pain	192
Pathology	Asphyxia	340
	Deaf	879

B. Feature Sets and Classifier Used

In this study, the performance of the proposed subband-TE and TECC feature sets is evaluated against the other state-of-the-art feature sets. The performance of the TECC feature set is compared against the other state-of-the-art cepstral features, namely, Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC), and Short-Time Cepstral Coefficients (STCC). Furthermore, subband-TE being a spectral representation, its performance is compared against the Mel Filterbank coefficients (MelFB), Linear Filterbank Coefficients (LinFB), and Short-Time Fourier Transform (STFT). The TECC, MFCC, LFCC, and STCC feature sets are of 120-D, 42-D, 120-D, and 120-D, respectively, and these feature sets include static, Δ, and $\Delta\Delta$ features. MelFBs and LinFBs are extracted by applying Mel filterbank and linear filterbank on STFT, respectively, using 40 subband filters in the corresponding filterbank. Furthermore, for fair comparison, subband-TE features also uses 40 subband filters. Thus, subband-TE, MelFB, and LinFB features are of 40-D, and STFT is represented using 257-D.

As this study focuses on handcrafted feature sets for infant cry analysis, we utilized the conventional state-of-the-art classifiers, namely, Gaussian Mixture Model (GMM)-based classifier and Support Vector Machine (SVM) [26]. GMM is utilized to learn the distribution of each class using a mixture of the Gaussian probability density functions (pdfs) represented by the mean, covariance, and weights. Then, test utterances are presented to the GMM of each class to finally estimate the log-likelihood ratio, which is used to predict the class of the test sample [26]. SVM is a non-probabilistic binary linear classifier which gives an optimal hyperplane for given labeled training data, and categorizes new examples [27].

Performance of the various systems are evaluated using two evaluation metrics, namely, $\%$ Equal Error Rate and $\%$ classification accuracy. The LLR scores obtained from CM system is used to compute EER. The EER is derived from detection error trade-off (DET) curve, which represents the performance on detection tasks that involve the trade-off of error types, namely, false alarm rate and miss rate [28]. The EER refers to the threshold at which both the error rates are equal. % classification accuracy is used to measure the number of samples classified correctly for a given class in $\%$. Furthermore, DET curves are also plotted for various classification system.

V. Experimental Results

A. Spectrographic Analysis

In Fig. 2, Panel-I and Panel-II represents the spectrographic analysis for randomly sampled normal and asphyxia cry signals, respectively. Fig. 2(a), Fig. 2(b), and Fig. 2(c) represents the STFT, MelFB, and subband-TE representations, respectively. It can be observed from Fig. 2(a) that there is difference in the pattern formed by F_0 and its harmonics for normal *vs.* asphyxia cry signals. These differences in the pattern is also visible for MelFB representation as shown in Fig. 2(b). However, these differences are more vivid for subband-TE representations as shown in Fig. 2(c). It might be because of the fact that TEO can accurately estimate the energy of the signal considering non-linear aspects of the speech production mechanism and also properties of airflow pattern in the vocal tract system [21], [23]. Furthermore, the results obtained using *10*-fold cross-validation also validates that the proposed TECC and subband-TE representations performs better over the other feature sets in this study.

TABLE II: Results in (% classification accuracy and % EER) using various cepstral feature sets using GMM and SVM as classifiers.

		MFCC	LFCC	STCC	TECC
GMM	Acc.	98.55	98.28	98.99	**99.12**
	EER	1.23	0.50	0.26	0.61
SVM	Acc.	88.11	80.18	87.84	86.56
	EER	12.72	18.78	13.84	12.57

B. Results

Cepstral representations are being common in speech signal processing applications, we performed the experiments using four cepstral feature sets, namely, MFCC, LFCC, STCC, and TECC. As the size of the dataset is relatively small, experiments are performed using *10*-fold cross-validation. The

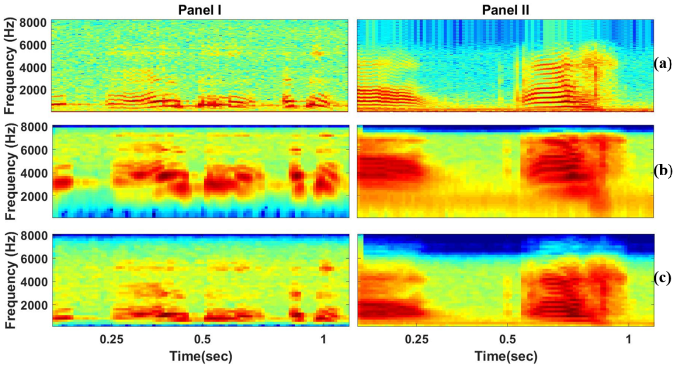

Fig. 2: Panel-I and Panel-II represents the spectrographic analysis for normal *vs.* asphyxia cry samples, respectively. Fig. 2(a), Fig. 2(b), and Fig. 2(c) represents the STFT, MelFB, and subband-TE representations, respectively.

TABLE III: Results in (% classification accuracy and % EER) for various spectral feature sets using GMM and SVM as classifiers.

		MelFB	LinFB	STFT	subband-TE
GMM	Acc.	98.99	98.77	98.59	**99.47**
	EER	1.5	0.70	1.6	**0.3678**
SVM	Acc.	88.15	87.80	78.06	**90.35**
	EER	10.49	10.40	19.41	**8.23**

dataset consists of the healthy and pathology class cry samples recorded with sampling rate of 22 kHz and 11 kHz, respectively. The experiments are performed using features extracted from the cry samples resampled to 16 kHz and results are reported in Table II. It can be observed that the proposed TECC feature set outperforms the other feature sets for both SVM and GMM classifiers. We utilized 512 Gaussian mixtures in the GMMs. Furthermore, experiments are extended with spectral feature sets, namely, subband-TE, MelFB, LinFB, and STFT. We utilized the spectral feature representations as it has low-dimensional representations than the cepstral features. It can be observed from Table III that the proposed subband-TE feature set outperforms the other feature sets for both SVM and GMM classifiers. Furthermore, all the spectral representations performs equally well as compared to their corresponding cepstral representations. However, subband-TE performs slightly better than its cepsral counterpart, i.e., TECC. Hence, it would be better to choose the spectral representations for this application.

Furthermore, DET curves are plotted for various spectral features as shown in Fig. 3. It can be observed that the proposed subband-TE representation performs better than all the other spectral representations for both the classifiers. The experiments are extended for varying number of Gaussian mixtures in GMM and results are obtained as shown in Table IV. It can be observed that the performance is improving as we increase the number of Gaussian mixtures in GMM from 64 to 512 and then it saturates, possibly due to the fact that a large number of 1024 mixtures is not required to model relatively

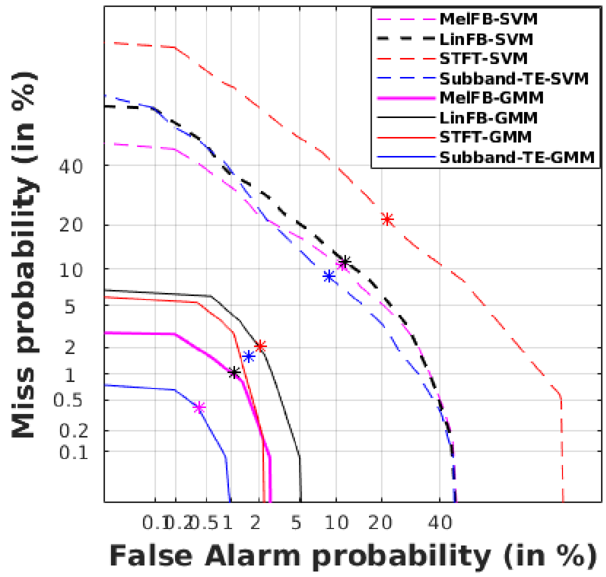

Fig. 3: DET plots for various feature sets using GMM and SVM as classifiers.

lesser duration of infant cry samples. Hence, we utilized 512 Gaussian mixtures in GMM for the remaining experiments. Furthermore, performance is also validated w.r.t. number of subband filters in the Gabor filterbank to extract the sunband-TE representations, and the results are reported in Table V. Fundamentally, TEO is developed for monocomponent signal. However, the speech signal occupies wide frequency range. To analyze the speech signal using TEO, we need to approximate it to monocomponent signal, and it is achieved using subband filtering. If we increase the number of subband filters in the filterbank then we obtain the better approximation to monocomponent signal. Hence, we validated the performance w.r.t. number of subband filters. It can be observed that the performance is almost constant w.r.t. number of subband filters in the filterbank and hence, we chose 40 number of filters in the filterbank as an optimal choice.

TABLE IV: Results (in % classification accuracy) w.r.t. number of mixtures

Mixtures	64	128	256	512	1024
Accuracy	98.72	98.94	99.16	**99.47**	99.47

TABLE V: Results in % classification accuracy (Acc) for various number of filters using GMM.

Filters	Acc.	Filters	Acc.	Filters	Acc.	Filters	Acc.
40	**99.47**	60	99.21	80	99.47	100	99.38
120	99.47	140	99.38	160	99.38	180	99.47

VI. Summary and Conclusions

In this study, we investigated the suitability of the spectral representations over cepstrals for infant cry analysis and classification. Because of the high pitch-source harmonics, spectral representations are more suitable for the normal *vs.* pathological infant cry classification. This theoretical assumption is validated using the experimental results. Furthermore, we exploited the capability of the TEO for accurately estimating the energies (especially approximated for the lower frequency regions). TEO being capable of better approximating the energies in low frequency regions, it is the suitable choice to extract information for pitch-source harmonics of infant cry, which is present at low as well as high frequency regions of the spectrogram.

VII. Acknowledgements

Authors would like to sincerely thank the doctors and organizers of the National Institute of Astrophysics and Optical Electronics, CONACYT Mexico for designing statistically meaningful Baby Chilanto database. The authors would also like to thank the Ministry of Electronics and Information Technology (MeitY), New Delhi, Govt. of India, for sponsoring a consortium project titled 'Speech Technologies in Indian Languages' under 'National Language Translation Mission (NLTM): BHASHINI', subtitled 'Building Assistive Speech Technologies for the Challenged' (Grant ID: 11(1)2022-HCC (TDIL)). We also thank the consortium leaders Prof. Hema A. Murthy, Prof. S. Umesh, and the authorities of DA-IICT Gandhinagar, India for their support and cooperation to carry out this research work.

References

[1] A. Chittora, "Crying for a reason: A signal processing based approach for infant cry analysis and classification," Ph.D. dissertation, Dhirubhai Ambani Institute of Information and Communication Technology, 2016.

[2] J. J. Engelsma, D. Deb, K. Cao, A. Bhatnagar, P. S. Sudhish, and A. K. Jain, "Infant-ID: Fingerprints for global good," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[3] H. A. Patil, "Infant identification from their cry," in *2009 Seventh International Conference on Advances in Pattern Recognition*, 2009, pp. 107–110.

[4] R. Colton and A. Steinschneider, "The cry characteristics of an infant who died of the sudden infant death syndrome," *Journal of Speech and Hearing Disorders*, vol. 46, no. 4, pp. 359–363, 1981.

[5] C. C. Onu, I. Udeogu, E. Ndiomu, U. Kengni, D. Precup, G. M. Sant'Anna, E. Alikor, and P. Opara, "Ubenwa: Cry-based diagnosis of birth asphyxia," *arXiv preprint arXiv:1711.06405*, 2017.

[6] C. C. Onu, J. Lebensold, W. L. Hamilton, and D. Precup, "Neural transfer learning for cry-based diagnosis of perinatal asphyxia," *ICLR Workshop*, 2019.

[7] K. Manickam and H. Li, "Complexity analysis of normal and deaf infant cry acoustic waves," in *Fourth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, 2005.

[8] O. Wasz-Höckert, T. Partanen, V. Vuorenkoski, K. Michelsson, and E. Valanne, "The identification of some specific meanings in infant vocalization," *Experientia*, vol. 20, no. 3, pp. 154–154, 1964.

[9] Q. Xie, R. K. Ward, and C. A. Laszlo, "Determining normal infants' level-of-distress from cry sounds," in *Proceedings of Canadian Conference on Electrical and Computer Engineering*. IEEE, 1993, pp. 1094–1096.

[10] H. A. Patil, ""cry baby": Using spectrographic analysis to assess neonatal health status from an infant's cry," in *Advances in speech recognition*. Springer, 2010, pp. 323–348.

[11] Y. Kheddache, C. Tadj *et al.*, "Characterization of pathologic cries of newborns based on fundamental frequency estimation," *Engineering*, vol. 5, no. 10, p. 272, 2013.

[12] H. F. Alaie, L. Abou-Abbas, and C. Tadj, "Cry-based infant pathology classification using gmms," *Speech Communication*, vol. 77, pp. 28–52, 2016.

[13] C. Ji, T. B. Mudiyanselage, Y. Gao, and Y. Pan, "A review of infant cry analysis and classification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2021, no. 1, pp. 1–17, 2021.

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

[15] D. Dimitriadis, P. Maragos, and A. Potamianos, "Auditory Teager energy cepstrum coefficients for robust speech recognition," in *INTERSPEECH*, Lisbon, Portugal, Sept. 2005, pp. 3013–3016.

[16] D. T. Grozdic and S. T. Jovicic, "Whispered speech recognition using deep denoising autoencoder and inverse filtering," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 25, no. 12, pp. 2313–2322, 2017.

[17] A. T. Patil, R. Acharya, P. A. Sai, and H. A. Patil, "Energy separation-based instantaneous frequency estimation for cochlear cepstral feature for replay spoof detection," *INTERSPEECH,Graz, Austria*, pp. 2898–2902, Sept. 2019.

[18] 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 *INTERSPEECH*, Stockholm, Sweden, Sept. 2017, pp. 12–16.

[19] M. R. Kamble and H. A. Patil, "Detection of replay spoof speech using teager energy feature cues," *Computer Speech & Language*, vol. 65, p. 101140, 2021.

[20] A. V. Oppenheim, R. W. Schafer, and T. Stockham, "Nonlinear filtering of multiplied and convolved signals," *IEEE Transactions on Audio and Electroacoustics*, vol. 16, no. 3, pp. 437–466, 1968.

[21] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Pearson Education, 3^{rd} edition, India, 2006.

[22] J. Lim, "Spectral root homomorphic deconvolution system," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 3, pp. 223–233, 1979.

[23] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, G. M. Hernández *et al.*, *Signals & systems*. Pearson Educación, 1997.

[24] H. M. 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.

[25] A. Rosales-Pérez, C. A. Reyes-García, J. A. Gonzalez, O. F. Reyes-Galaviz, H. J. Escalante, and S. Orlandi, "Classifying infant cry patterns by the genetic selection of a fuzzy model," *Biomedical Signal Processing and Control*, vol. 17, pp. 38–46, 2015.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[28] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *EUROSPEECH*, Rhodes, Greece, 1997, pp. 1895–1898.