Teager Energy Cepstral Coefficients for Classification of Normal *vs.* Whisper Speech

Kuldeep Khoria, Madhu R. Kamble, Hemant A. Patil
Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat, India.
{kuldeep_khoria, madhu_kamble, hemant_patil}@daiict.ac.in

***Abstract*—The whispered speech is quite different from natural speech in the context of nature, acoustic characteristics, and generation mechanism. In order to improve the robustness of Automatic Speech Recognition (ASR) system, it is very important to analyze the mismatched training and testing situations and propose a robust acoustic features to enhance the whisper recognition. In this paper we propose to use Teager Energy Cepstral Coefficients (TECC) which uses Teager Energy Operator (TEO) for estimating "true" total energy of the signal, i.e., the sum of kinetic and potential energies which is contradictory to the traditional signal energy approximation, which only takes kinetic energy into account, i.e., L^2 norm of the signal. In this study, experiments are performed on wTIMIT and CHAINS corpus. For wTIMIT corpus, frame-level accuracy of 92.22 % is obtained and for CHAINS corpus, it is 95.61 %. We have also estimated the performance measure of the classifier by using Matthew Correlation Coefficient (MCC), F-measure, and J-statistics. Furthermore, experiments are performed by considering latency period from a practical deployment viewpoint, and the trade-off between latency period *vs.* accuracy is discussed for both the corpora.**

***Index Terms*—Whispered Speech Recognition (WSR), Teager Energy Operator, Equal Error Rate (EER), Latency.**

I. Introduction

Whisper speech detection has become a research topic of interest [1], [2]. The differences in whispered *vs.* normally phonated speech are primarily due to the noisy structure, lower Signal-to-Noise-ratio (SNR), absence of glottal vibrations, shift in formant structures, etc. [3]. In Automatic Speaker Recognition (ASR) systems, the training and testing are performed on normal and corresponding whisper, i.e., mismatch speech dataset, and hence, the performance degrades. Several approaches have been proposed to attenuate the mismatch through feature transformations [4], model adaptation [5]–[8], or using alternative sensing technologies, such as throat microphone [9].

Whispered Speech Recognition (WSR) is an active research field but, one of the major limitation is that the corpus is not systematically and suitably collected. There are few publicly available databases for parallel normal and corresponding whispered speech collected for different languages, such as English [10], [11], Mandarin language [12], Japanese [5], and Serbian language. However, there is limited dataset with there corresponding transcription and also the vocabulary is limited in size. One of the first experiments on automatic whisper recognition is reported in [5]. The key goal was to recognize the whispered speech on cell phones in real world conditions. Using MFCC-HMM system, they analyzed different mismatched train/test scenarios by taking three speech modes, namely, whisper, low-voice speech, and neutral speech. Severe degradation in ASR was reflected due to use of mismatch data. However, there was outstanding result when ASR model is trained on whisper (whisper speech model), and also it was working well enough for testing with either type of speech.

Unlike normal speech, whisper speech does not contain fundamental frequency (Fo) due to the absence of voice harmonic distortion, and formant shifting in the lower frequency regions [5], [13]. In [13], [14], it has been observed that normal and whisper speeches have different formant characteristics, where vowels of Serbian and English language were used. It was observed that the formant frequency F_1 for whisper speech is greater than that of normal speech for both female and male speakers. This characteristics can be used as a main attribute to classify the normal and whisper speech. The same study explored that formant bandwidths for whisper vowels has a general expansion than that of vowels.

To improve the performance of ASR system, recently various approaches have been proposed for the conversion of whispered speech to normal speech with the aim of improving speech intelligibility, and naturalness [15]–[20]. As the use of voice assistants, and Text-to-Speech (TTS) systems is becoming more common and hence, the need for the speaker to interact with such systems privately is also increasing simultaneously. In such scenarios, a user may wish to whisper to the device, and would also expect a response in a whispered voice, as is the case with the recently released version of *Amazon Alexa* [21]. To further improve robustness of these ASR systems, some pre-processing can be performed by developing clusters of normal, and whisper speech so that they can be identified beforehand, and further processing can be done accordingly, in particular, ASR of whispered speech [22]. Due to this reason, the classification of these two types of speeches becomes an important part of ASR systems.

II. Acoustic Features Used

A. LFCC vs. MFCC

While extracting MFCC or LFCC feature sets, the speech signal is windowed and DFT is computed for each frame to get the Short-Time Fourier Transform (STFT), $X(n, \omega_m)$. The energy in STFT is weighted by each Mel scale filter frequency response, $U_l(\omega)$, to get the l^{th} energy coefficient, i.e.,

$$E_{mel}[n,l] = \frac{1}{A_l}\sum_{k=L_l}^{V_l} |U_l(\omega_m)X(n,\omega_m)|^2. \tag{1}$$

The real cepstrum C_{mel} associated with the $E_{mel}(n,k)$ is referred to as MFCC:

$$C_{mel}[n,p] = \frac{1}{R}\sum_{k=0}^{R-1} log(E_{mel}(n,k))cos(\frac{2\pi}{R}kp), \tag{2}$$

where R is the number of subband filters. The transformation in eq. (2) is also known as Discrete Cosine Transform (DCT). In this paper, we have considered MFCC as the baseline feature set to compare the result (because MFCC is one of the most successful feature representation for several speech signal processing applications) [23], [24]. Both MFCC, and LFCC use similar algorithm for feature extraction except the type of frequency response used to obtain the weighted sum from the spectrum. In general, Mel scale gives more significance to the lower frequency regions, and less significance to the higher frequency regions [25]. This arrangement suggests that the MFCC fails to extract effective spectral characteristics at the high frequency range. Both MFCC and LFCC feature sets use triangular-shaped filters in order to obtain the subband filtered components. This means that the features which can retain both low frequency, and high frequency characteristics could be effective for classification of whisper *vs.* normal speech.

B. TECC

The basics of TECC feature extraction process is similar to the MFCC, however, it is having one major difference in terms of estimating the energy. For TECC feature set, the nonlinear Teager Energy Operator (TEO) is employed that estimates the instantaneous Teager energy instead of standard energy (Squared Energy Operator (SEO) that employ L^2 norm of a signal) [26]. For a monocomponent discrete-time signal, $x[n]$, TEO, ($\Psi_d\{\cdot\}$), is defined as [27]:

$$E_n = \Psi_d\{x[n]\} = x^2[n] - x[n-1]x[n+1], \tag{3}$$

where E_n gives the running estimate of signal's energy which is under consideration. The TEO performs well on the monocomponent signal and cannot be applied directly on speech signal (as speech signal itself is the combination of several monocomponent signals). Hence, the speech signal is is passed through bandpass filter to obtain *'N'* narrowband filtered signals, and then the TEO is applied on the i^{th} narrowband filtered signal, i.e., $\Psi_d\{x_i[n]\}$.

The nonlinear modeling mechanism of the speech production is the main motivation behind using TEO instead of standard SEO [28]. In conventional linear phonic theory, it is assumed that airflow propagates as a plane wave from the vocal tract system. However, this assumption may not hold for real speech signal as it is produced due to nonlinear nature of vortex flow interactions [29]. Since, the whispered speech are nonlinear and contains extreme turbulent airflow, applying TEO for find the cepstral characteristics can help us to capture the difference in airflow pattern of the speech production. The TEO incorporates both amplitude and frequency information and computes the ’true’ total source energy of a speech signal [30] along with improving time-frequency components of rapid energy changes [26]. In addition, it also improves the formant representation information in the feature vectors. The functional block diagram of TECC feature extraction process is shown in Fig. 1. Here, the input speech signal is passed through the pre-emphasis filter, and further applied to the Mel-spaced Gabor filterbank to obtain subband filtered signals. The signal output of subband filter are then applied to the TEO block to estimate the instantaneous Teager energy profiles of each subband signals. Furthermore, these Teager energy profiles are allowed to pass from the frame-blocking, and averaged with a frame length of 20 ms and overlapping of 10 ms followed by compressing the data by logarithm operation. To obtain a low-dimensional representation that has compact energy, Discrete Cosine Transform (DCT) is applied along with Cepstral Mean Normalization (CMN) (also known as Cepstral Mean Subtraction (CMS)) to reduce the channel mismatch/distortion conditions [31]. Finally, retained few DCT coefficients to get Teager Energy Cepstral Coefficients (TECC) which are appended along with their Δ and $\Delta\Delta$ features to obtain higher-dimensional feature vector [32].

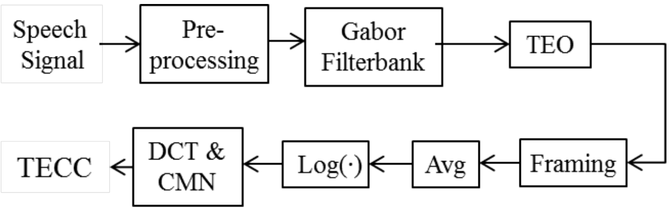

Fig. 1. Block diagram of TECC feature extraction. After [32].

In addition, we compared the spectral energy densities (as shown Fig. 2) obtained from the traditional Short-Time Fourier Transform (STFT), and Teager energy-based approach for the both corpora, namely, wTIMIT and CHAINS. In particular, Panel I and Panel II in Fig. 2 shows the natural and corresponding whisper speech for wTIMIT corpus and Panel III and Panel IV shows for natural and whisper speech for CHAINS corpus. The spectral energies for the time-domain signal for traditional STFT and Teager energy-based approach are shown in Fig. 2 (b) and Fig. 2 (c), respectively. It can be observed that the energy density obtained from the Teager energy-based approach preserves much more information in low as well as in high frequency regions as compared to the traditional spectrogram. In particular, the formants are well preserved for the natural speech (refer Panel I (b)) which are not visible for the traditional spectrogram for wTIMIT corpus. In case of CHAINS corpus, it is observed that many of the higher frequency information are not preserved when estimated from the traditional spectrogram whereas the Teager energy-based approach do carry the information in the higher frequency regions. This spectral energy obtained from the Teager energy-

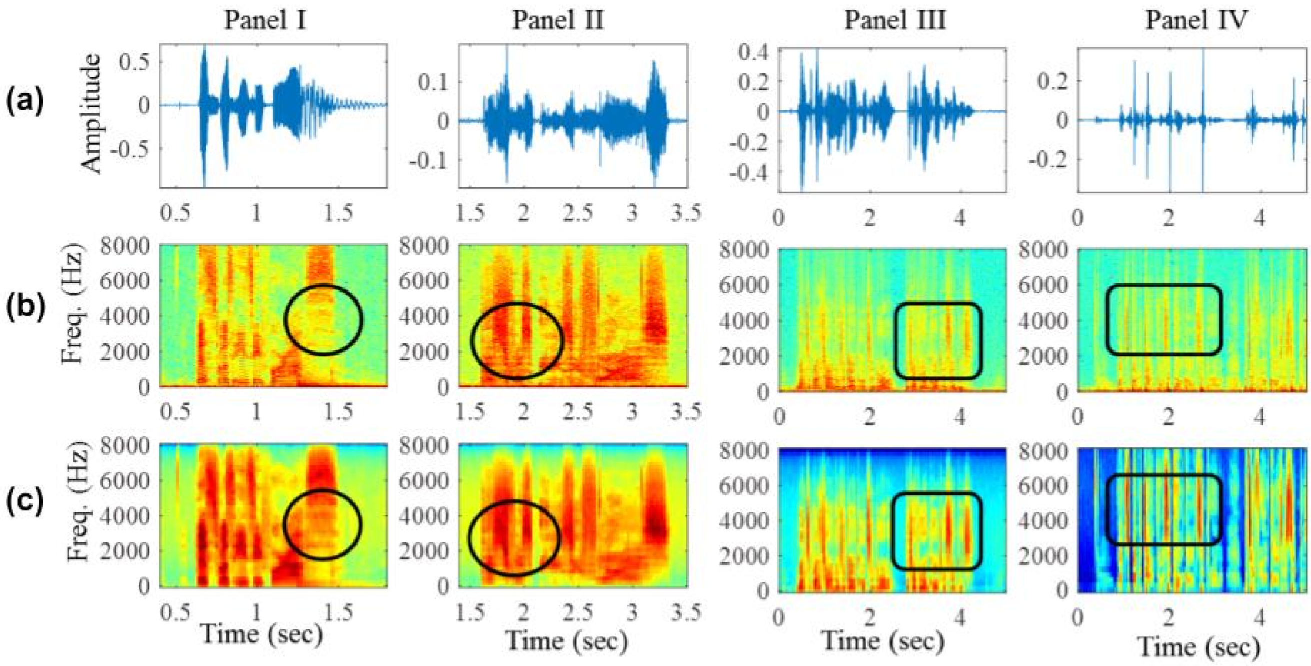

Fig. 2. Panel I and Panel II are the natural and corresponding whisper speech from the wTIMIT corpus, Panel III and Panel IV are the natural and corresponding whisper speech from the CHAINS corpus. (a) Time-domain speech signal, (b) traditional STFT spectrogram, and (c) spectral energy density obtained from Teager energy-based approach. The discriminative regions are indicated by circle and box for corresponding wTIMIT and CHAINS corpora.

based method indeed help to classify the whisper speech from its natural counterpart.

III. Experimental Setup

A. Corpora Used

We performed the experiments on two corpora, namely, wTIMIT and CHAINS. The wTIMIT corpus was collected in two phases, the first phase was recorded in Singapore, and the second phase was recorded in the USA [6]. In this paper, we used only the USA recorded data. The sampling frequency of data is set as 44.1 kHz, and all the recordings were done in clean acoustic environment. For training, we have 9219 and 11325 for normal and whisper utterances whereas 412 whisper utterances, and 727 normal utterances are used for testing. The CHAINS corpus is designed to characterize speakers as individuals. The corpus contains the recordings of 36 speakers (20 male and 16 female) in two different sessions with a time separation of about two months. For training, we considered 1036 normal and whisper utterances, and 296 normal and whisper utterances for testing.

B. Feature Extraction Parameters

All the utterances during feature extraction process were first resampled to 16 kHz from 44.1 kHz. This is done primarily so as to reduce the number of samples thereby saving computational cost. The process of frame-blocking is carried out by taking a window length of 20 ms with an overlap of 10 ms. We have considered *39*-D feature vector extracted from 40 number of subband filters in a filterbank for MFCC, LFCC, and TECC feature set. We have taken logarithm, and then DCT to obtain static coefficients appending along with Δ and $\Delta\ \Delta$ to obtain *39*-D feature vector.

C. Pattern Classifier

In this study, Gaussian Mixture Model (GMM) is used as a two-class pattern classifier, where the two classes corresponds to the speech samples of the normal *vs.* whispered speech. The individual GMM is trained for each class using LFCC, MFCC, and TECC feature set. The Expectation Maximization (EM) algorithm is exploited in GMM to find out the maximum likelihood estimation (MLE) parameters for the given data. The log-likelihood (llk) score $s(X)$ for each test sample is estimated using the trained GMM as in:

$$s(X) = llk(X|\lambda_w) - llk(X|\lambda_n), \tag{4}$$

where λ_w, and λ_n represents the GMM trained on whisper, and normal speech samples, respectively, and X represents a new testing sample. The scores obtained helps to classify whether the unknown sample belongs to the natural or whisper. We have taken 512 Gaussian mixtures for this work. The robustness and feature discrimination power of our proposed feature set is also evaluated using Matthew correlation coefficient (MCC), F-measure, and J-statistics. Furthermore, we used a standard evaluation metric, i.e., Equal Error Rate (EER) which is indicated on the Detection Error Trade-off (DET) curve for the whisper detection system [33]. The DET curve is used to study the performance of the SSD system. When operating point in the DET curve of False Acceptance Rate (FAR), and False Rejection Rate (FRR) or miss probability is *equal*, then it is referred to as EER.

IV. Experimental Results

The experiments are performed on wTIMIT and CHAINS corpus with TECC feature set are shown in Table I. We observed the effect of two different frequency scales, namely, linear and Mel scale in the Gabor filterbank in order to obtain the subband filtered signals according to the center frequencies. When the features are extracted using linear frequency scale, the accuracy of the whisper speech detection was not high as when the features were extracted using Mel frequency scale for both the corpora. Along with observing the effect of frequency scale, we also observed the effect of applying the CMN technique for both the corpora. It can be observed from the Table I that the accuracy obtained from the Mel frequency scale along with CMN technique gave better accuracy of 92.22 % on wTIMIT, and 95.61 % on CHAINS corpus, respectively.

TABLE I
Accuracy (in %) for TECC feature set on wTIMIT and CHAINS corpora

Corpus	CMN	Frequency Scale	Accuracy (%)
wTIMIT	×	Linear	84.65
	✓	Linear	90.93
	×	Mel	82.78
	✓	Mel	**92.22**
CHAINS	✓	Linear	86.88
	✓	Mel	**95.61**

In addition, we also observe the feature discrimination power using F measure, J-statistic, and MCC as shown in Table II. We observe that the TECC feature set has high values for all the measures as compared to the other feature sets and thus, it is more discriminative to classify natural *vs.* whisper speech signals for both corpus.

TABLE II
Analysis of Feature Discrimination Power using F-measure, J-statistic, and MCC

Corpus	Feature Sets	MCC	F-measure	J-measure
wTIMIT	LFCC	0.73	0.89	0.75
	MFCC	0.61	0.83	0.63
	TECC	**0.83**	**0.93**	**0.86**
CHAINS	LFCC	0.67	0.83	0.67
	MFCC	0.43	0.64	0.44
	TECC	**0.92**	**0.95**	**0.91**

TABLE III
Results in terms of EER and Accuracy in (%)

Feature Sets	EER		Accuracy	
	wTIMIT	CHAINS	wTIMIT	CHAINS
LFCC	12.59	16.05	86.82	83.97
MFCC	17.37	5.97	80.12	94.06
TECC	**6.69**	**4.46**	**92.22**	**95.61**

Furthermore, the performance evaluation metric is computed in terms of EER for all the feature sets. It can be observed from Table III that we obtain low EER for the TECC feature set compared to THE other MFCC and LFCC feature sets. For wTIMIT corpus, the low EER with TECC feature set is 6.69 % and for CHAINS corpus, it is 4.46 %. The performance evaluation is also shown in Fig. 3 by the DET curves for MFCC, LFCC, and TECC feature sets. It can be observed that the miss probability of MFCC, and LFCC is very high for the given FAR, which is not a good case for whisper detection system. There is significant decrease in miss probability for TECC feature set for wTIMIT as shown in Figure 3(a). We observe similar pattern of results on CHAINS corpus, as shown in Figure 3(b). However, the TECC and MFCC feature sets have low miss probability, and LFCC feature set has very high miss probability. We also analyzed the trade-off between latency period *vs.* accuracy (as shown in Fig. 4) for (a) wTIMIT, and (b) CHAINS corpora. Here, latency period refers to the duration between the speech utterance produced to the system, and response from the system in terms % of accuracy i.e. number of frames considered to classify the utterance. In the other words, if a system gives better accuracy for lower latency periods, then it means that this system would not be waiting for the entire utterance to judge whether the utterance is natural or whisper. Instead, lower levels of latency with higher accuracy would ensure that faster classification of natural and whisper utterances. In this graph, we considered frame-level accuracy. It can be clearly observed from the graph that the accuracy of TECC feature set increases continuously as the latency is increased. This behavior is expected because if the number of frames taking part in accuracy calculation increases, then the average value of accuracy tend to increase.

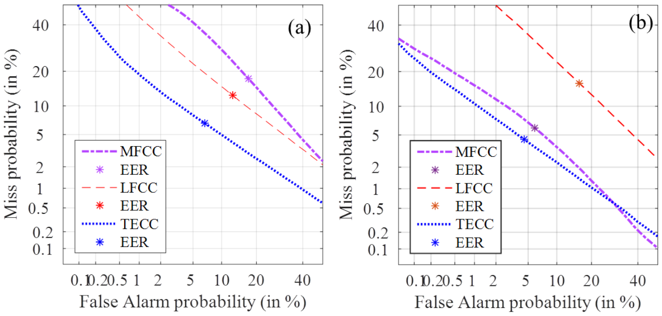

Fig. 3. DET curve for TECC, MFCC and, LFCC feature set for (a) wTIMIT, and (b) CHAINS corpus.

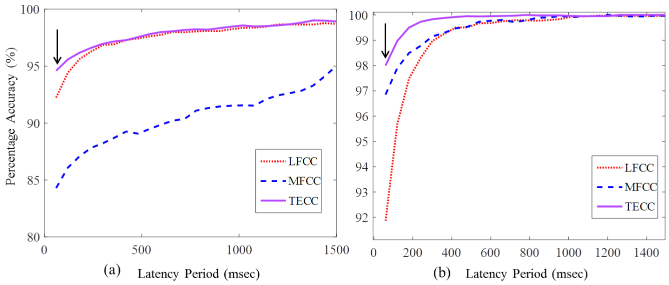

Fig. 4. Accuracy (in %) *vs.* latency period for TECC, MFCC, and LFCC feature set for (a) wTIMIT, and (b) CHAINS corpus

V. Summary and Conclusions

In this work, we explored TECC, MFCC, and LFCC feature set for normal $vs.$ whisper speech classification. As whispered speech contains nonlinear and extremely turbulent airflow, the feature representation should incorporates both amplitude and frequency information of the signal. Hence, estimating the "true" total energy of the signal instead of estimating only kinetic energy into account of the signal. By listening to the speech samples of natural and whisper speech, it has been observed that the initial and the end portion of the utterance consists of silence regions. These silence regions produces ambiguity to the classification architecture. We tried to eliminate these regions. However, it is not the straightforward to remove such silent regions for the whispered signal because the amplitude of acoustic noise is much higher than the amplitude of the whisper component. Hence, noise cannot be directly removed from the whispered speech. Thus, developing efficient Voice Activity Detection (VAD) algorithms in whispered speech is also a potential area of research. With this development, we can significantly improve the frame-level classification accuracy for the whispered speech and also higher accuracy at low latency period.

References

[1] B. R. Marković, J. Galić, and M. Mijić, "Application of teager energy operator on linear and mel scales for whispered speech recognition," *Archives of Acoustics*, vol. 43, 2018.

[2] . T. Grozdić and S. T. Jovičić, "Whispered speech recognition using deep denoising autoencoder and inverse filtering," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2313–2322, 2017.

[3] R. W. Morris, "Enhancement and recognition of whispered speech," Ph.D. dissertation, Georgia Institute of Technology, 2003.

[4] C.-Y. Yang, G. Brown, L. Lu, J. Yamagishi, and S. King, "Noise-robust whispered speech recognition using a non-audible-murmur microphone with vts compensation," in *IEEE 8th International Symposium on Chinese Spoken Language Processing*, Hong Kong, China, 2012, pp. 220–223.

[5] T. Ito, K. Takeda, and F. Itakura, "Analysis and recognition of whispered speech," *Speech Communication*, vol. 45, no. 2, pp. 139–152, 2005.

[6] B. P. Lim, "Computational differences between whispered and non-whispered speech," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2011.

[7] A. Mathur, S. M. Reddy, and R. M. Hegde, "Significance of parametric spectral ratio methods in detection and recognition of whispered speech," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 157, 2012.

[8] S. Ghaffarzadegan, H. Boşil, and J. H. Hansen, "Generative modeling of pseudo-target domain adaptation samples for whispered speech recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. South Brisbane, Queensland, Australia: IEEE, 2015, pp. 5024–5028.

[9] S.-C. Jou, T. Schultz, and A. Waibel, "Adaptation for soft whisper recognition using a throat microphone," in *International Conference on Spoken Language Processing*, Jeju Island, Korea, pp. 1493–1496.

[10] C. Zhang and J. H. Hansen, "Whisper-island detection based on unsupervised segmentation with entropy-based speech feature processing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 883–894, 2010.

[11] S. Ghaffarzadegan, H. Bořil, and J. H. Hansen, "Ut-vocal effort ii: Analysis and constrained-lexicon recognition of whispered speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, 2014, pp. 2544–2548.

[12] P. X. Lee, D. Wee, H. S. Y. Toh, B. P. Lim, N. F. Chen, and B. Ma, "A whispered mandarin corpus for speech technology applications," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[13] S. T. Jovičić, "Formant feature differences between whispered and voiced sustained vowels," *Acta Acustica united with Acustica*, vol. 84, no. 4, pp. 739–743, 1998.

[14] J. B. Wilson and J. D. Mosko, "A comparative analysis of whispered and normally phonated speech using an LPC-10 vocoder," Rome Air Development Center Giffis AFB NY, Tech. Rep., 1985.

[15] C. Huang, X. Y. Tao, L. Tao, J. Zhou, and H. B. Wang, "Reconstruction of whisper in chinese by modified melp," in *2012 7th International Conference on Computer Science & Education (ICCSE)*, Melbourne, Australia., 2012, pp. 349–353.

[16] I. V. Mcloughlin, H. R. Sharifzadeh, S. L. Tan, J. Li, and Y. Song, "Reconstruction of phonated speech from whispers using formant-derived plausible pitch modulation," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 6, no. 4, pp. 1–21, 2015.

[17] R. W. Morris and M. A. Clements, "Reconstruction of speech from whispers," *Medical Engineering & Physics*, vol. 24, no. 7-8, pp. 515–520, 2002.

[18] H. R. Sharifzadeh, I. V. McLoughlin, and F. Ahmadi, "Reconstruction of normal sounding speech for laryngectomy patients through a modified celp codec," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 10, pp. 2448–2458, 2010.

[19] V.-A. Tran, G. Bailly, H. Loevenbruck, and T. Toda, "Improvement to a nam-captured whisper-to-speech system," *Speech Communication*, vol. 52, no. 4, pp. 314–326, 2010.

[20] T. Toda, M. Nakagiri, and K. Shikano, "Statistical voice conversion techniques for body-conducted unvoiced speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2505–2517, 2012.

[21] M. Cotescu, T. Drugman, G. Huybrechts, J. Lorenzo-Trueba, and A. Moinet, "Voice conversion for whispered speech synthesis," *IEEE Signal Processing Letters*, vol. 27, no. 01, pp. 186–190, 2019.

[22] J. H. Hansen, C. Zhang, and X. Fan, "Speech processing for robust speaker recognition: Analysis and advancements for whispered speech," in *Forensic Speaker Recognition, Neustein and Patil (Eds.)*. Springer, 2011, pp. 253–272.

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

[24] D. A. Reynolds, "Speaker identification and verification using gaussian mixture speaker models," *Speech Communication*, vol. 17, no. 1-2, pp. 91–108, 1995.

[25] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of mfcc," *Journal of Computer Science and Technology*, vol. 16, no. 6, pp. 582–589, 2001.

[26] J. F. Kaiser, "Some useful properties of teager's energy operators," in *1993 IEEE international conference on acoustics, speech, and signal processing*, vol. 3. IEEE, 1993, pp. 149–152.

[27] Kaiser, James F, "On a simple algorithm to calculate the energy of a signal," in *IEEE ICASSP*, Albuquerque, New Mexico, USA, 1990, pp. 381–384.

[28] P. Maragos *et al.*, "Speech nonlinearities, modulations, and energy operators," in *IEEE ICASSP*, Toronto, Canada, 1991, pp. 421–424.

[29] 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.

[30] D. Dimitrios, M. Petros, and P. Alexandros, "Auditory Teager energy cepstrum coefficients for robust speech recognition." in *INTERSPEECH*, Lisboa, Portugal, 2005, pp. 3013–3016.

[31] S. Molau, F. Hilger, and H. Ney, "Feature space normalization in adverse acoustic conditions," in *IEEE ICASSP*, Hong Kong, China, 2003, pp. I–656–659–I.

[32] M. R. Kamble and H. A. Patil, "Analysis of reverberation via Teager energy features for replay spoof speech detection," in *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, Brighton, UK, 2019, pp. 2607–2611.

[33] A. Martin *et al.*, "The DET curve in assessment of decision task performance," in *EUROSPEECH*, Rhodes, Greece, 1997, pp. 1895–1898.