An Efficient Speaker Recognition System by Employing BWT and ELM

A. Indumathi[1] and **E Chandra**[2]

Submitted in April, 2016; Accepted in July, 2016

Abstract – Speaker recognition system has gained substantial research interest, owing to security enforcement in many applications. Mostly, the speaker recognition system is employed for achieving access control. Hence, a speaker recognition system must be capable of achieving greater recognition accuracy rates irrespective of the noise presence. This paper presents a novel speaker recognition system which tends to suppress the noise before the process of feature extraction. This idea improves the recognition accuracy of the system. Additionally, the proposed work can manage the noise mismatch between both the training and testing phase. Bionic Wavelet Transform (BWT) is applied to suppress the noise and the cleansed signal is obtained. This is followed by extracting Mean Hilbert envelope Coefficients (MHEC) and Power Normalized Cepstral Coefficients from the cleansed signal. Finally, the classification is achieved by Extreme Learning Machine (ELM). From the performance evaluation, it is evident that the proposed work shows convincing recognition accuracy rates, Signal to Noise ratio (SNR) and Mean Square Error (MSE).

Index Terms – Speaker recognition system, noise suppression, feature extraction, classification.

NOMENCLATURE

BWT-Bionic Wavelet Transform,*MHEC*-Mean Hilbert envelope Coefficients, ELM- Extreme Learning Machine, SNR-Signal to Noise ratio, MSE-Mean Square Error, MFCC -Mel Frequency Cepstral Coefficients, PLP-Perceptual Linear Prediction

1.0 INTRODUCTION

The process of communication relies on two entities namely speaker and listener [1]. On the whole, a message is transferred by means of words. However on keen observation, the language of the speech, emotion and the identity of speaker can be gained. The objective of a speaker recognition system is to extract features from the speech signal and to figure out the identity of the speaker [2]. Generally, a speaker recognition system involves two important processes, which are speaker verification and identification [3,4]. Speaker verification verifies the identity of the speaker by utilizing a sample speech. Speaker identification intends to detect the speaker from a set of speech samples. Speaker recognition system is mostly employed for security based applications such as access control, authentication, law enforcement and personalization [5].Many such speaker recognition systems require clean speech signal for proving high recognition accuracy have been presented. However, the issue being caused by noise is not addressed in most of the works.Most of the real time applications suffer fromexternal noise,which has a serious impact over the quality of the system [6,7].

The features being extracted from the noisy speech will turn dissimilar to that of the training dataset. This paves way for higher misclassification rates. Besides this, many existing works provide solutions for noise removal [8]; however, it is not mandatory that the training and testing samples of speech should possess the same type of noise.

This paper strives to present a novel speaker recognition system that overthrows the noise by means of bionic wavelet transform. The features are then extracted from the cleansed speech signal and finally classified. The Mean Hilbert envelope Coefficients (MHEC) and Power Normalized Cepstral Coefficients (PNCC) are extracted from the clean speech signal and fed into the Extreme Learning Machine (ELM). The entire work is decomposed into three key phases and they are noise suppression, feature extraction and classification. The major contributions of this paper are listed below.

This work can effectively deal with noisy speech signals and thus the quality of the system is improved.

As the features are extracted from the cleansed speech signal, the recognition accuracy of the work is enhanced.

The remainder of this paper is organized as follows. The review of literature is presented in section 2. The proposed approach is presented in section 3. Section 4 evaluates the performance of the proposed approach. Finally, the concluding remarks are drawn in section 5.

2.0 REVIEW OF LITERATURE

This section intends to review the related literature with respect to noise suppression, feature extraction and classification.

2.1 Noise suppression

A stereo based block-wise linear compensation approach is proposed in [9], which dealt with babble, white and office noise. The work presented in [10] identifies speaker by utilizing spectral features. The spectral features are extracted by latent prosody analysis. In [11], a method to deal with stationary noise is presented, by utilizing short time Fourier transform and Ephraim-Malah estimation. This work reduces

[1], Dept.of Computer Science, Dr SNS RCAS
[2]Bharathiar University, Coimbatore, Tamil Nadu
[1] endhumathi@gmail.com and [2] crcspeech@gmail.com.

the spectral coefficients and thus the noise is suppressed. The work proposed in [12] tends to suppress noise by the transformation of the noise affected signal to the wavelet domain and thereby the local maxima coefficients are conserved. Level dependent threshold is introduced in [13], in order to eliminate coloured noise. In [15], the bionic wavelet transformation is improvised by clubbing the wavelet denoisin approach, in order to construct wavelet thresholding scheme. The exploitation of wavelet filters through multistage convolution withthe help of reverse biorthogonal wavelets in both high and low pass frequency bands.

However, it is not certain that the noise present in the training sample have to exist in the test speech signal. Though the existing works assume that both phases suffer from the same kind of noise and this reduces the recognition accuracy of the system. The proposed work aims to suppress the noise irrespective of the noise type and thus can serve its purpose effectively.

2.2 Feature extraction

Mel Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction (PLP) are the predominantly used features for the process of speech recognition [15,16]. The major drawback of MFCC is that the background noise can bring in the issue of disparity between the training and testing phases. The main factors for this issue are given below. The spectrum being assessed by MFCC is susceptible to noise and channel distortion [17]. Besides this, the acoustic model of MFCC is not found to be accurate for speaker recognition. Realizing the above mentioned points, the proposed work strives to utilize a feature which does not suffer from the background noise.All the stated drawbacks are overcome by Mean Hilbert Envelope Coefficients (MHEC). The performance of MHEC is proven to be the best under noisy environment and noise mismatch between training and testing phase.

Power Normalized Cepstral Coefficients (PNCC) is recently proposed and is proven to be better than Zero Crossing Peak Amplitude (ZCPA), Relative Spectral Transform – Perceptual Linear Prediction (RASTA-PLP), Perceptual Minimum Variance Distortionless Response (PMVDR) and so on [18-20]. PNCC proves its efficiency with different kinds of noise in both training and testing phase. However, it is a universal fact that hybridization of different algorithms improves the recognition accuracy even more. Taking the aforementioned statement into account, this work combines the two different techniques namely MHEC and PNCC, in order to combat against noise mismatch and to achieve higher recognition accuracy rate.

2.3 Classification

Classification plays a vital role in speaker recognition. Mostly employed classifier for speaker recognition system is Support Vector Machine (SVM) [21,22]. Some of the other classifiers are k-Nearest Neighbour (k-NN) and decision trees [23,24]. Besides this, certain generative models such as Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) are also employed. This work proposes to exploit ELM as the classifier, owing to its effectiveness, simplicity and speed.

3. Proposed approach

The objective of this section is to describe the functionality of the proposed speaker recognition system. The proposed work is decomposed into three building blocks, which are noise suppression, feature extraction and classification. The task of noise suppression is achieved by bionic wavelet transform. The process of feature extraction aims to excerpt MHEC and PNCC from the cleansed speech signal. This step is followed by the classification, which achieves the task of speaker recognition. The forthcoming subsections present the description of all the phases and the overall schematic diagram is presented in figure 1.

3.1 Noise suppression

A speech signal carries both essential and unnecessary details of information. This unnecessary detail of a signal can be denoted as noise. Thus, it is mandatory to suppress noise, which in turn improves the quality of the signal. However, noise suppression is a separate area of research and this work shows a small concern towards it. The proposed work employs bionic wavelet transform for noise suppression.

3.1.1 Bionic Wavelet Transform (BWT)

Basically, BWT follows the principle of Morlet wavelet, which is exclusively designed with respect to the human voiced signal [14]. The wavelet transform contains the coefficients of the signal $x(t)$, with regard to $h_{\alpha,\tau}(t)$ and all these are the elements of $h(t)$. This $h(t)$ is the mother wavelet and it is represented by

$$h_{\alpha,\tau}(t) = \frac{1}{\sqrt{|\alpha|}} h\left(\frac{t-\tau}{\alpha}\right)(1)$$

The wavelet transform coefficients are obtained by the product of $x(t)$ and the basis functions,

$$W_x(\tau,\alpha) \leq x(t);\ h_{\alpha,\tau}(t) \geq \frac{1}{\sqrt{|\alpha|}} \int x(t) h^* \left(\frac{t-\tau}{\alpha}\right) dt(2)$$

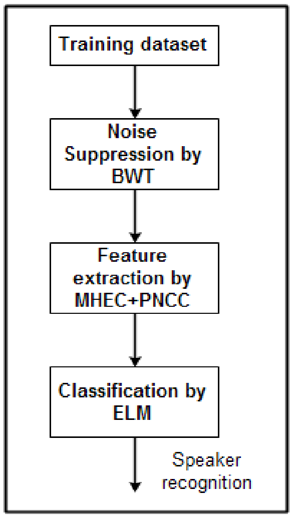

Fig 1: Overall schematic diagram of the proposed system

The basic idea of BWT is to substitute the constant quality factor with the changing adaptive quality factor [25]. This kind of substitution can be made by altering the mother function of the wavelet transform. The fluctuating $h(t)$ is denoted as

$$h(t) = \hat{h}(t) \exp(j2\pi f_0 t) \quad (3)$$

In the above equation, f_0 is the centre frequency of $h(t)$ and $\hat{h}(t)$ is the envelope function. T-value is employed in the BWT mother function and is represented in the below given equation.

$$h_T(t) = \frac{1}{T} \hat{h}\left(\frac{t}{T}\right) \exp(j2\pi f_0 t) \quad (4)$$

The BWT is represented by the following equations.

$$bwt_x(\tau, \alpha) = \frac{1}{\sqrt{|\alpha|}} \int x(t)\, h_T^{*}\left(\frac{t-\tau}{\alpha}\right) dt \quad (5)$$

This can be written as

$$bwt_x(\tau, \alpha) = \frac{1}{T\sqrt{|\alpha|}} \int x(t)\, \hat{h}^{*}\left(\frac{t-\tau}{\alpha T}\right) \times \exp\left(-j2\pi f_0 t \left(\frac{t-\tau}{\alpha}\right)\right) dt$$

(6)

From the above equations, it can be observed that the BWT mother function's amplitude and time spread depend on the value of T. The value of T can be generated on the basis of the idea of Yao and Zhang and is defined by

$$T(\tau + \Delta\tau) = \left(1 - \widehat{G1} \frac{bwt_s}{bwt_s + |bwt_x(\tau,\alpha)|}\right)^{-1} \times \left(1 + \widehat{G2} \left|\frac{\partial bwt_x(\tau,\alpha)}{\partial t}\right|\right)^{-1}$$

(7)

In the above equation, $\widehat{G1}$, $\widehat{G2}$ and bwt_s are constants, $bwt_x(\tau, \alpha)$ is the BWT coefficient at the time τ and scale α. Thus, it could be noted that the T function brings in adaptability to the BWT. The derivation of T function can be referred from [26,27]. The BWT coefficients are based on the WT coefficients and this can be represented by

$$bwt_x(\tau, \alpha; T) = k \times WT_x(\tau, \alpha) \quad (8)$$

The value of k depends on the value of T.

As already stated that the BWT is based on Morlet function, it is used as the mother function. The real morlet function is denoted by

$$h(t) = e^{-\left(\frac{t}{T_0}\right)^2} \quad (9)$$

The value of k is determined by eqn.10 as given in [28].

$$\frac{\int_{-\infty}^{+\infty} e^{-t^2} dt}{\sqrt{\left(\frac{T}{T_0}\right)^2 + 1}} \approx \frac{1.7725}{\sqrt{\left(\frac{T}{T_0}\right)^2 + 1}} \quad (10)$$

The noise suppression can be achieved by two levels of thresholds namely hard and soft thresholding techniques [29,30]. The degree of adaptability is based on the value of T function and so, T function is incorporated in determining the threshold. Besides this as per eqn.8, it is advisable to take least values for BWT coefficients, when the scales of decomposition are high. Thus, the T-function can be employed for lower scales of decomposition. The threshold can be calculated by

$$th = \frac{\sigma}{\sum_i \alpha_i T_{fs(i)}} \sqrt{2 \log_2 N} \quad (11)$$

The value of α_i is chosen by the trial and error method and it is more optimal to have decreasing function. After performing the operation of thresholding, the BWT coefficients are divided by the k factor and then inverse transform is performed. This rebuilds the signal to arrive at the cleansed signal. The next step is to extract features from the cleansed signal and is described below.

3.2 Feature extraction

Feature extraction is the most important phase of a speaker recognition system. This phase shows a great impact over the final classification stage. The better the choice of feature extractor, the greatest is the recognition accuracy rate. Thus, this work shows keen attention towards the

choice of feature extractors. After exhaustive study, we come to a conclusion that it would be optimal to choose feature extractors that can cope up with the noise mismatch during the training and testing phase. Understanding the benefits of hybridization, this work proposes to combine two feature extractors such as MHEC and PNCC and explained in the following subsections.

3.2.1 MHEC

The MHEC is recently proposed in [31], which effectively manages the noise mismatch in the training and testing stages. The cleansed signal $s(t)$ is passed for feature extraction, which is then divided into 24 bands with the help of a Gammatone filterbank [32]. The centre frequencies of the filterbank are placed on the rectangular bandwidth scale from 300 to 3400 Hz. The process involved in MHEC are presented in the below given equations.

$$s_a(t,j) = s(t,j) + i\hat{s}(t,j) \quad (12)$$

In the above equation, $\hat{s}(t,j)$ is the Hilbert transform of $s(t,j)$ and i is the imaginary part. The temporal envelope or Hilbert envelope $e_s(t,j)$can be c (13)computed by

$$e_s(t,j) = s^2(t,j) + \hat{s}^2(t,j)(13)$$

The duplicate high frequency components are suppressed by smoothening the temporal envelope with a low pass filter. The cut off frequency is set as 20 Hz.

$$e_{sn}(t,j) = (1-\alpha)e_s(t,j) + \alpha e_{sn}(t-1,j) \quad (14)$$

Where α is the smoothening factor and is inversely proportional to the cut-off frequency.

The resultant smoothened temporal envelope is then blocked into frames with duration25 ms and the skip rate is fixed as 10 ms. Each frame is then applied with a hamming window, so as to reduce the edge discontinuities. The temporal envelope amplitude in frame p is computed by

$$s(p,j) = \frac{1}{N}\sum_{t=0}^{N-1} w(t)\, e_{sn}(t,j) \quad (15)$$

$w(t)$ represents the hamming window and N is the size of the frame.

Natural log is applied to compress the spectral parameters. Besides this, Discrete Cosine Transform is applied to convert the spectrum to cepstrum and to decorrelate the feature dimensions. The outcome of this process is 36 dimensional feature vectors.

3.2.2 PNCC

The attractive features of PNCC are listed below. The introduction of its power-law nonlinearity, which is the substitute for log nonlinearity in MFCC paves way for robustness. The building blocks of PNCC are pre-processing, temporal integration, asymmetric noise suppression, temporal masking, spectral weight smoothening and mean power normalization.

Initially, a pre-emphasis filter which is of the kind $H(z) = 1 - 0.97z^{-1}$ is applied. This step is followed by the application of Short-Time Fourier Transform (STFT) by hamming windows. The factor $\hat{Q}[m,l]$ is only utilized for noise assessment and compensation. This can also be employed to alter the information with respect to the short time power estimates. This step is followed by the specifically designed asymmetric filter for noise suppression. Temporal masking is achieved by moving peak for every frequency channel and the power is suppressed, if at all it falls below the mask. The masked signals can serve its purpose in a better way, when it is exposed to reverberant environment. The outcome of the channels are then smoothened in spectral weight smoothening. Mean power normalization is incorporated into the system, so as to reduce the impact of amplitude scaling.

The noise mismatch is effectively handled by both MHEC and PNCC. Thus, efficient feature vectors are obtained. The so formed feature vectors are then fed into the classifier, in order to identify the speaker.

3.3 Classification by ELM

ELM is one of the effective classifiers with faster learning capability. The weights between the input and the hidden layer are allocated randomly. This can be achieved by a generalized inverse operation of the hidden layer output matrix.

Consider a training dataset (a_i, z_i) and $i = 1,2,..N$ and $a_i \in F^R$; $z_i \in F^q$. In this case, the ELM training is done by the following steps.

1. Allocate values to the lower layer weight matrix in a random fashion of the range [-1,1], such that $W \in F^{R.v}$, and v is the count of hidden units.

2.The hidden layer outcome is computed for each training sample a_i, such that $hl_i = \sigma(W^T\ a_i)$ in which σ is the sigmoid function.

3. The weight of output layer O is computed by $O = (HH^T)^{-1}HT^T$. where $H = [h1, h2, \dots hn]$ and $T = [t1, t2, .. tn]$

The random weight assignment is done irrespective of the training dataset and thus the new data can also be well-generalized. In this work, the ELM is trained with different voice samples, so as to classify between different speakers. The ELM's output for the test speech signal presents a k dimensional vector, which presents the top ranked ELM score from the trained dataset. The sample with maximum score is recognized as the speaker.

Thus, all the stages involved in the proposed work are described. The next section intends to evaluate the performance of the proposed approach.

4.0PERFORMANCE EVALUATION

The performance of the proposed work is analysed by exploiting KING Dataset [33]. The attributes of the dataset are given below.

Table 1: Details of king Dataset

Details	King dataset

Speaker count	51
Session count	10
Speech kind	Photograph description
Mic	Dual
Channel	Dual
Rate of sampling	8 kHz
Digital quantization	16 bits

While performing the experimental analysis, certain noises such as Babble noise, train, airport and car noise are included in the speech signal at different scales such as -5, 0, 5, 10 and 15 dB. All these noises are downloaded from the Noisex-92 database [32].

The signal quality is measured in terms of Signal-to-Noise Ratio (SNR) and Mean Square Error (MSE).

$$SNR = \frac{f^2}{n^2} \quad (16)$$

$$MSE = \frac{1}{N}\Sigma\left(f - \hat{f}\right)^2 \quad (17)$$

In the above equations, f is the clean signal and $\hat{f}$ is noised eliminated signal and n is the noise signal.

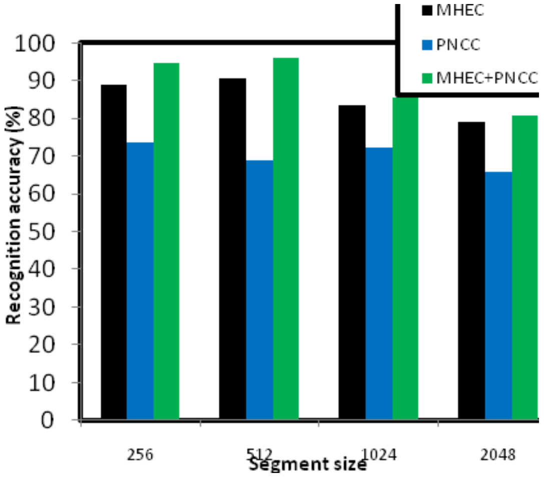

Fig 2: Recognition accuracy with k-means classification

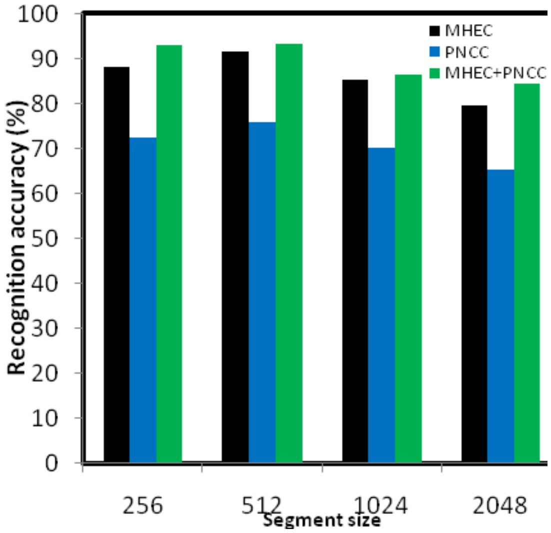

Fig 3: Recognition accuracy with SVM classification

The proposed approach is compared with two standard noise suppression methods such as spectral subtraction (SS) and Iterative Wiener Filtering (IWF) [32,34]. Besides this, we prove the effectiveness of the hybridization by carrying out the feature extractors individually. The overlap percentage is fixed as 60% and the results are obtained. Additionally, the comparison is made with respect to the classifier also. The proposed work is compared with k-means and SVM. Figure2,3 and 4 depicts the recognition accuracy rate ofthe system with respect to the k-means, SVM and ELM classification respectively.

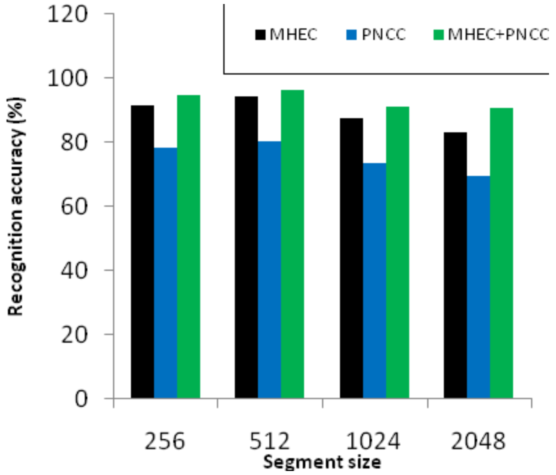

Fig 4: Recognition accuracy with ELM classification

On analysing the above graphs, it could be noted that the performance of the hybrid approach shows consistently good results. Besides this, the recognition accuracy rate of ELM is greater which when compared to k-means and SVM. The next part strives to compare the experimental resultswith respect to SNR and MSE and the graphical results are presented below.

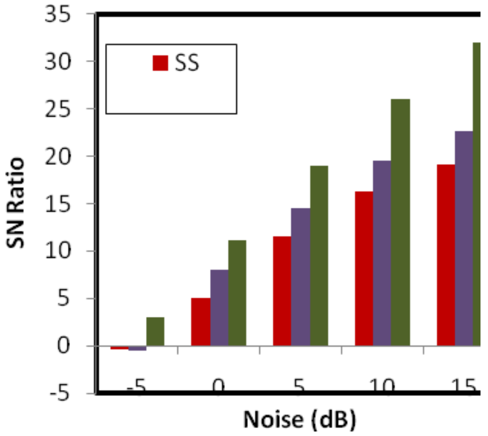

Fig 5: Signal to Noise Ratio

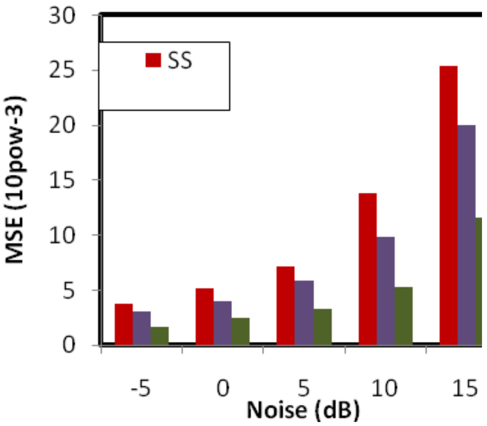

Fig 6: Mean Square Error

From the obtained experimental results, the performance of the proposed work is found to be satisfactory in terms of SNR, MSE and recognition accuracy rates. We found that the system performs well with the segment size of 256 and 512. Thus, the proposed work serves its purpose with greater accuracy rates.

5.0 CONCLUSION

This paper strives to present a novel speaker recognition system, which can effectively manage noise mismatch during training and testing phase. Initially, the noise is suppressed with the bionic wavelet transform with adaptive filtering technique. The obtained cleansed signal is then passed through the step of feature extraction. In this step, the features such as MHEC and PNCC are clubbed together, in order to inherit the merits of both the techniques. Finally, ELM is employed as the classifier to recognize between several speakers. On experimental analysis, the proposed work shows greater recognition accuracy rate and least MSE.

REFERENCES

[1]. Rupayan Das and Pradip K.Das, “Design and Implementation of Monophones and Triphones –Based Speech Recognition Systems for voice Activated Telephoney”,BIJIT-BVICAM's International Journal of Information Technology,Vol.5,No.1,2013.

[2]. Ruchi Chaudhary, “Short –Term Spectral Feature Extraction and Their Fusion in Text Independent Speaker Recognition: A Review”,BIJIT-BVICAM's International Journal of Information Technology,Vol.5,No.2,2013.

[3]. H. Beigi, Fundamentals of Speaker Recognition,Springer, New York, 2011.

[4]. A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14(1) (2004) 4–20.

[5]. Douglas A. Reynolds, "An Overview of Automatic Speaker Recognition Technology", in Proc. of International Conference on Acoustics, Speech and Signal Processing, 2002.

[6]. T. Kinnunen, R. Saeidi, F. Sedlak, K.A. Lee, J. Sandberg, M. Hansson-Sandsten, H.Li, Low-variance multitaper MFCC feature: a case study in robust speaker verification, IEEE Trans. Audio Speech Lang. Process. 20(1) (2012) 1990–2001.

[7]. S.K. Nemala, K. Patil, M. Elhilali, A multistream feature framework based on bandpass modulation filtering for robust speech recognition, IEEE Trans. Audio Speech Lang. Process. 21(2) (2013) 416–426.

[8]. Y. Wang, M.J.F. Gales, Speaker and noise factorization for robust speech recog-nition, IEEE Trans. Audio Speech Lang. Process. 20(7) (2012) 2149–2158.

[9]. L. Deng, A. Acero, M. Plumpe, X. Huang, Large vocabulary speech recognition under adverse acoustic environments, in: Proceedings of the Sixth Interna-tional Conference on Spoken Language Processing, Beijing, China, October 2000, pp.806–809.

[10]. Y.F. Liao, Z.H. Chen, Y.T. Juang, Latent prosody analysis for robust speaker iden-tification, IEEE Trans. Audio Speech Lang. Process. 15(6) (2007) 1870–1883.

[11]. Z. Brajevic, A. Petosic, Signal denoising using STFT with Bayesprediction and Ephraim–Malahestimation, in: Proceedings of the 54th International Symposium ELMAR, Zadar, Croatia, September2012, pp.183–186.

[12]. S. Mallat, W.L. Hwang, Singularity detection and processing with wavelets, IEEE Trans. Inf. Theory 38(2) (1992) 617–643.

[13]. I.M. Johnstone, B.W. Silverman, Wavelet threshold estimators for data with cor-related noise, J. R. Stat. Soc. 59(2) (1997) 319–351.

[14]. J. Yao, Y.T. Zhang, Bionic wavelet transform: a new time–frequency method based on an auditory model, IEEE Trans. Biomed. Eng. 48(8) (2001) 856–863.

[15]. Chanwoo Kim and Richard M. Stern, Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition, in Proc. of International Conference on Acoustics, Speech and Signal Processing, Mar.25-30, pp. 4101-4104,2012.

[16]. H. Hermansky, "Perceptual linear prediction analysis of speech," J. Acoust. Soc. Am., vol. 87, no. 4, pp. 1738–1752, Apr. 1990.

[17].] J. Alam, T. Kinnunen, P. Kenny, P. Ouellet, and D. O Shaughnessy, "Multi-taper MFCC features for speaker verification using i-vectors," in Proc. IEEE ASRU, Hawaii, HI, Dec. 2011, pp. 547–552.

[18]. D.-S. Kim, S.-Y. Lee, and R. M. Kil, "Auditory processing of speech signals for robust speech recognition in real-world noisy environments," IEEE Trans. Speech and Audio Processing, vol. 7, no. 1, pp. 55–69, 1999.

[19]. H. Hermansky and N. Morgan, "RASTA processing of speech," IEEE. Trans. Speech Audio Process., vol. 2, no. 4, pp. 578–589, Oct. 1994.

[20]. U. H. Yapanel and J. H. L. Hansen, "A new perceptually motivated MVDR-based acoustic front-end (PMVDR) for robust automatic speech recognition," Speech Communication, vol. 50, no. 2, pp. 142–152, Feb. 2008.

[21]. H. Hu, M.-X. Xu, and W. Wu, "GMM supervector based SVM with spectral features for speech emotion recognition," in Proceedings of IEEE ICASSP 2007, vol. 4. IEEE, 2007, pp. IV–413.

[22]. T. L. Nwe, N. T. Hieu, and D. K. Limbu, "Bhattacharyya distance based emotional dissimilarity measure for emotion classification," in Proceedings of IEEE ICASSP 2013. IEEE, 2013, pp. 7512–7516.

[23]. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," in Proceedings of Interspeech, 2009, pp.320–323.

[24]. Y. Kim and E. Mower Provost, "Emotion classification via utterance-level dynamics: A pattern-based approach to characterizing affective expressions," in Proceedings of IEEE ICASSP 2013. IEEE, 2013.

[25].] J. Yao and Y. T. Zhang, "Bionic wavelet transform: a new timefrequency method based on an auditory model," IEEE Transactions on Biomedical Engineering, vol. 48, no. 8, pp. 856–863, 2001.

[26]. J. Yao and Y. T. Zhang, "Cochlear is an inhomogeneous, active and nonlinear model," in Proceedings of the 1st Joint Meeting of BMES & IEEE/EMBS, p. 1031, Atlanta, Ga, USA, October 1999.

[27]. J. Yao and Y. T. Zhang, "From otoacoustic emission modeling to bionic wavelet transform," in Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, pp. 314–316, Chicago, Ill, USA, July 2000.

[28]. X. Yuan, "Auditory model-based bionic wavelet transform for speech enhancement," M.S. thesis, Speech and Signal Processing Laboratory, Marquette University, Milwaukee, Wis, USA, 2003.

[29]. D. L. Donoho, "De-noising by soft-thresholding," IEEE Transactions on Information Theory, vol. 41, no. 3, pp. 613–627, 1995.

[30]. D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," Journal of the American Statistical Association, vol. 90, no. 432, pp. 1200–1224, 1995.

[31]. Seyed Omid Sadjadi, Taufiq Hasan, and John H.L. Hansen, "Mean Hilbert Envelope Coefficients (MHEC) for Robust Speaker Recognition", in Interspeech, pp. 1696-1699, Sep. 2012 .

[32]. R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, "Complex sounds and auditory images," in Auditory Physiology and Perception, Y. Cazals, L. Demany, and K. Horner, Eds. Oxford: Pergamon Press, 1992, pp. 429–446.

[33]. A. Higgins, D. Vermilyea, KING speaker verification, http://catalog.ldc.upenn.edu/ldc95s22, 1995.

[34]. P.S.Banerjee,Baisakhi Chakraborty and Jaya Banerjee, "Feature Extraction of Voice Segments using Ceptural Analysis for Voice Regeneration", BIJIT-BVICAM's International Journal of Information Technology, Vol.7, No.2, 2015.