ISSN: 2229-6948(ONLINE)
DOI: 10.21917/ijct.2016.0187

EVALUATION OF SOUND CLASSIFICATION USING MODIFIED CLASSIFIER AND SPEECH ENHANCEMENT USING ICA ALGORITHM FOR HEARING AID APPLICATION

N. Shanmugapriya[1] and E. Chandra[2]
[1]Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts and Science, India
E-mail: spriyanatrajan@gmail.com
[2]Department of Computer Science, Bharathiar University, India
E-mail: crcspeech@gmail.com

Abstract

Hearing aid users are exposed to diversified vocal scenarios. The necessity for sound classification algorithms becomes a vital factor to yield good listening experience. In this work, an approach is proposed to improve the speech quality in the hearing aids based on Independent Component Analysis (ICA) algorithm with modified speech signal classification methods. The proposed algorithm has better results on speech intelligibility than other existing algorithm and this result has been proved by the intelligibility experiments. The ICA algorithm and modified Bayesian with Adaptive Neural Fuzzy Interference System (ANFIS) is to effectiveness of the strategies of speech quality, thus this classification increases noise resistance of the new speech processing algorithm that proposed in this present work. This proposed work indicates that the new Modified classifier can be feasible in hearing aid applications.

Keywords:

Independent Component Analysis (ICA), Speech Intelligibility, Bayesian Modified with Adaptive Neural Fuzzy Interference System (ANFIS)

1. INTRODUCTION

In this field of research, speech intelligibility and quality, which are extremely vital for hearing loss people and it can be improved considerably through enhancement [1]. Hearing aids supported with the assistance of speech enhancement algorithms effectively assist the hearing loss people for the purpose of understanding the speech even in different noisy environments.

Speech enhancement and noise reduction methods have been widely used in noisy environment applications like hearing aids devices, these methods have also been used in other different applications such as mobile phone systems and speech coding systems. The main aim of speech enhancement is to improve the performance of speech communication systems in noisy environments. Speech enhancement can be applied to a speech recognition system or a mobile radio communication system, a set of low quality recordings or to improve the performance of aids for the hearing impaired.

The interference source may be a wide-band noise in the form of a white or colour noises, a periodic signal such as in room reverberations, hum noise, or it can take the form of diminishing noise. The speech signal may be simultaneously attacked by one or more noise source. Even though speech enhancement has drawn severe research [2] and algorithms stimulated from various characteristics have been developed, it is still an unresolved problem [3] since there are no accurate models for both speech and noise [2]. Algorithms in accordance with multiple microphones [4] and single microphone has also been effective in achieving certain measure of speech enhancement [5-7]. Although single channel noise removal techniques such as spectral subtraction, Kalman filtering, and Wiener filtering exist, they often yield limited noise removal results. Multi-microphone based techniques such as beam forming [8] and Independent Component Analysis (ICA) [9] typically offer superior performance when compared to the single microphone situation.

Recently, the intention of multichannel speech enhancement has been re-modified, with the intention that noise reduction can be achieved without dereverberating speech. Contrary to the beam forming schemes, the knowledge regarding the geometry of the microphone array is not necessary, and the optimal filter completely depends on the second-order statistics of the noisy signal. In case of [10], the authors formulated the most familiar schemes of multichannel noise reduction in accordance with the linear filtering. In this scheme, the noise-free speech is completely estimated through a linear transformation of the observation vector. The most uncomplicated approach is to diminish the Mean Square Error (MSE) among the noise-free and filtered speech signals at a given microphones, which gives way for a multichannel version of the classical Wiener filter. In this scenario, certain noise is trimmed down at the cost of the amplified speech distortion, however this cannot explicitly manage the trade-off among these quantities. Here this proposed work used noise removal algorithm ICA for good performance [11] and the modified Bayesian with ANFIS for speech signal classification in hearing aid applications.

In this proposed work present a common framework to study the most important noise reduction and classification in the hearing aid application. The use of multiple microphones can help in minimizing speech distortion while having a good amount of noise reduction at the same time. This work is structured as follows. Section 2 describes the related work and section 3 describes the proposed methodologies. In section 4 explains experimental results and finally give conclusions in Section 5.

2. RELATED WORK

Veisi, et al. [12] formulated a Hidden Markov Model (HMM)-based minimum MSE speech enhancement schemes in Mel-frequency domain and a Parallel Cepstral and Spectral (PCS) modelling. Both Mel-Frequency Spectral (MFS) and Mel-Frequency Cepstral (MFC) characteristics are completely investigated and experimented for the purpose of effective speech enhancement. In order to fairly accurate clean speech waveform in the scenario of a noisy signal, an inversion from the Mel-frequency domain to the spectral domain is necessary which initiates distortion artifacts in the spectrum assessment and the filtering. With the intention of diminishing the corrupting consequences of the inversion, the PCS modelling is formulated. This approach carries out concurrent modelling in both cepstral and magnitude spectral domains. Together with the spectrum estimator, magnitude spectrum, log-magnitude spectrum and power spectrum estimators are also investigated and assessed in the HMM-based speech enhancement framework.

Van den Bogaert, et al [13] assessed speech enhancement in binaural multi micro phone hearing aids through noise reduction schemes in accordance with the Multichannel Wiener Filter (MWF) and the MWF with partial noise estimate (MWF-N). Both these schemes are exclusively developed for the purpose of combining noise reduction with the protection of binaural cues. Objective and perceptual assessments were carried out with different speech-in-multitalker-babble constructions in two different acoustic atmospheres. The foremost conclusions are: (a) A bilateral MWF with ideal voice activity detection equals or works better than a bilateral adaptive directional microphone in terms of speech enhancement at the same time preserving the binaural cues of the speech component. (b) A considerable gain in speech enhancement is found when sending one contra lateral microphone signal to the MWF active at the ipsilateral hearing aid. Adding together a second contra lateral microphone showed a considerable development at some point in the objective evaluations however not in the subset of scenarios tested for the period of the perceptual evaluations. (c) Adding the partial noise estimate to the MWF, done to enhance the spatial awareness of the hearing aid user, decreases the quantity of speech enhancement in a partial way. In certain circumstances the MWF-N even do better than the MWF possibly because of an improved spatial release from masking.

Digital hearing aid users frequently find fault of complexity in understanding speech in the existence of background noise. In order to enhance speech perception in a noisy atmosphere, several speech enhancement schemes have been applied in digital hearing aids. Here, a speech enhancement scheme by means of modified spectral subtraction and companding is formulated for digital hearing aids. Lee, et al [14] fine-tuned the biases of the estimated noise spectrum, in accordance with a subtraction factor, to diminish the residual noise. Companding was applied to the channel of the formant frequency in accordance with the speech presence indicator to improve the formant. Noise suppression was attained at the same time retaining weak speech constituents and avoiding the residual noise phenomena. Objective and subjective assessment under a variety of environmental conditions confirmed the enhancement because of this scheme.

Hearing aid users typically experience feedback which roots "howling" and limits the maximum stable gain. Shusina, et al. [15] formulated the direct scheme of adaptive feedback cancellation is extensively utilized for the purpose of mitigating feedback, on the other hand it is less efficient for high forward path gains because of the intrinsic bias in the feedback path estimate. This bias can be considerably reduced at the cost of artificial delays which can adequately initiate pre-echo and "comb filter" effects. The direct scheme also tends to terminate tonal audio signals, for instance, alarms and music. Here, the authors formulated using a system identification scheme in closed-loop for unbiased feedback cancellation which does not have the negative features manifested by the direct scheme, however makes use of an identification signal. The two-stage scheme which makes use of two adaptive filters, one to recognize the complete closed-loop, and another to extract the feedback path response was the preferred unbiased scheme because of its low computation and better feedback identification particularly during "howling".

Binaural hearing aids are permitting hearing impaired persons getting more information from surroundings (for instance, the route of incident, stereo sound effect, etc.) than monaural, and the binaural noise reduction schemes turn out to be essential for this category of assistive hearing devices. Li, et al. [16] intended to measure the powers of speech enhancement procedures to the target source binaural auditory localization.

The studies of all the above mentioned methodologies to use multiple microphones with different speech enhancement and noise reduction techniques in order to better deal with the fundamental problems in Speech intelligibility and quality. Here this proposed work is used to ICA for good performance in noise reduction and for better classification results, used Modified Bayesian-ANFIS, and the experimental results for both subjective and objective tests prove the dominance of the proposed schemes in the speech analysis, particularly for non-stationary (Speech) noises.

3. PROPOSED METHODOLOGY

The block diagram of the Proposed Methodology in speech processing is shown in Fig.1.

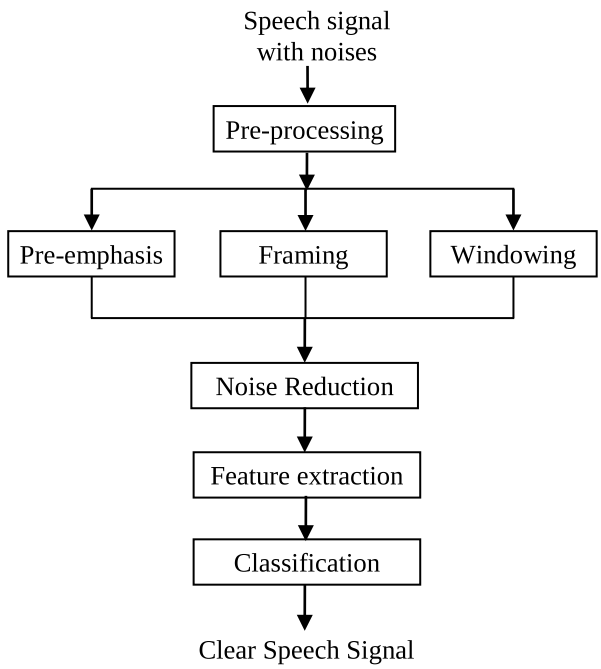

Fig.1. Block Diagram for Speech Processing

3.1 PREPROCESSING METHOD FOR SPEECH SIGNAL ENHANCEMENT

Signal pre-processing is performed to enhance the speech signals in the hearing aid application. In order to enhance the desired audio signal, a pre-processing step is applied to speech enhancement process. This enhancement may consist of a pre-processing for enhances a signal from a certain direction and a noise reduction process that is used to reduces the noise based signal properties. The most widely using pre-processing steps in speech signal processing are, 1) Pre-emphasis, 2) Framing and 3) Windowing, which are described as follows:

3.1.1 Pre-emphasis using FIR Filtering:

The speech signal $s(t)$ with noise signal $n(t)$ is considered as a $x(t)$, which is put through a first order Finite Impulse Response (FIR) filter to spectrally level the speech signal to make it less vulnerable to finite accuracy effects later in signal processing in the hearing aid application.

Let us consider $\{x(t)\}$ be the impulse response of a casual FIR filter of length $T + 1$ and the transfer function of this filter is,

$$X(z) = \sum_{t=0}^{T} x[t] z^{-t} \tag{1}$$

The corresponding frequency response of the signal is given by,

$$X\left(e^{j\omega}\right) = \sum_{t=0}^{T} x[t] e^{-jn\omega} \tag{2}$$

where, T is the order of the filter. In FIR filters, exact linearity of the phase is achieved, if the speech signal frequency response of the filter is expressible in the form,

$$X\left(e^{j\omega}\right) = \bar{X}(\omega) e^{j\phi(\omega)} \tag{3}$$

where, $\phi(\omega) = \alpha\omega + \beta$, ω is the frequency point of the speech signal and $\bar{X}(\omega)$ is a real and even function of ω. The magnitude and phase of the above function are respectively,

$$\left|X\left(e^{j\omega}\right)\right| = \left|\bar{X}(\omega)\right| \tag{4}$$

And,

$$\arg X\left(e^{j\omega}\right) = \begin{cases} \alpha\omega + \beta \text{ for } \bar{X}(\omega) \geq 0 \\ \alpha\omega + \beta - \Pi \text{ for } \bar{X} < 0 \end{cases}. \tag{5}$$

$\bar{X}(\omega)$ is known as the zero phase frequency response of the given input speech signal of the filter to distinguish it from the magnitude response $X(e^{j\omega})$. To simplify, $\underline{X}(\omega)$ = zero phase frequency response. It should be clear from the environment whether X is function of z, $e^{j\omega}$ or ω i.e., whether the transfer functions and the frequency response is considered. The zero phase frequency response of the filter may take both speech and noise signal values, whereas the magnitude response is strictly without noise in the speech signal.

3.1.2 Framing:

In speech processing it is often advantageous to divide the signal into frames to achieve stationarity, because the speech signal is non-stationarity. In this framing step the pre-emphasised signal $s(\tilde{t})$ is converted into speech samples. The first frame consists of X speech samples. The second frame begins Y samples after the first frame and overlaps through X-Y samples. This procedure persists until the entire speech is accounted for within one or more frames. The values of X and Y have to be selected such that $X \geq Y$, to make sure that the adjacent frames overlap. If Y is chosen to be greater than X, some of the speech signal would be lost and signal estimates may include a noisy component.

3.1.3 Windowing:

The next step in pre-processing is to perform windowing on individual frames in an attempt to diminish the signal discontinuities during the beginning and at the end of each frame. This speech signal enhancement approach uses the window to narrow the signal to zero during the beginning and the end of each frame. Consider the window as $w(t)$, $0 \leq t \leq Y - 1$ and the result of windowing is the signal given by Eq.(6).

$$x_t(\tilde{t}) = x_t(t) * w(t), \qquad 0 \leq t \leq X - 1 \tag{6}$$

A typical window used in the speech signal enhancement is Hamming window, which has the form given by Eq.(7).

$$w(t) = 0.54 - 0.46\cos(2t/(X-1)), \quad 0 \leq t \leq X - 1 \tag{7}$$

After the pre-processing, speech signal $s(t)$ in the hearing aid application also have the some background noise signals, such as $n(t)$.

3.2 NOISE REDUCTION IN HEARING AIDS

Background noise tends to decrease the speech intelligibility especially for people suffering from hearing loss. To reduction of these noise signals and extract the clean speech signal, this speech enhancement method used ICA. The ICA uses multiple microphone recordings, while the noise reduction is applied to a single acoustic signal.

3.2.1 Independent Component Analysis (ICA):

Before applying speech signals to ICA, it is usually very useful to do some pre-processing. The pre-processing techniques are centring and whitening, that make the problem of ICA estimation simpler and better conditioned [17].

3.2.2 Centring for ICA:

The necessary pre-processing is to centre the mixed signal q.Centring means, subtract it mean vector $x = E\{q\}$ from q, so as to make q as zero mean variable. This pre- processing is made solely to simplify the ICA algorithms: It does not mean that the mean could not be estimated. After estimating the mixing matrix A with centred data, this work can complete the estimation by adding the mean vector of s back to centred estimates of s. The mean vector is given by, $A^{-1}x$, where x is the mean that was subtracted in the pre-processing [17].

3.2.3 Whitening for ICA:

Another pre-processing strategy in ICA is to initially whiten the examined variables. Following the cantering the data and before the application of the ICA algorithm, this work used to transform the observed vector q linearly, so that obtained a new vector $\tilde{q}$ which is whitened. Its component is uncorrelated and their variances are equal to unity. The covariances are equal to unity. The covariance matrix $\tilde{a}$ equivalent to the identity matrix.

$$E\left(\tilde{q}\tilde{q}^{T}\right)=I \tag{8}$$

The whitening alteration is possible all the time. One most significant scheme for the purpose of whitening is to make use of the Eigen-Value Decomposition (EVD) of the covariance matrix $E\left(\tilde{q}\tilde{q}^{T}\right)=EDE^{T}$, where E stands for the orthogonal matrix of the eigenvectors of the $E\left(\tilde{q}\tilde{q}^{T}\right)$ and D stands for the diagonal matrix of its Eigen-values. Whitening completely transforms the mixing matrix into a new one $\tilde{A}$.

$$\tilde{q}=ED^{-\frac{1}{2}}E^{T}As=\tilde{A}s \tag{9}$$

The whitening completely diminishes the amount of parameters to be estimated. Rather than having to estimate n^2 parameter that is the original matrix A, this work needs to estimate orthogonal mixing matrix $\tilde{A}$. Whitening is very simple and standard procedure for ICA algorithm, it is efficiently reduce the complexity [17].

Independent Component Analysis (ICA) is an adaptive, linear representation method. When applied to speech frames, ICA provides a linear representation that maximizes the statistical independence of its coefficients, and therefore finds the directions with respect to which the coefficients are as lightly distributed as possible. The ICA [18] basis function matrix denoted by A, and the filter matrix by W. Independent component analysis aims to recover a set of unknown mutually independent sources signals from there linear mixtures are observed without knowing the mixing coefficients.

Let the received background noise corrupted speech signal in the hearing aid application be $x(t) = s(t) + n(t)$, where m is the discrete time index, $s(t)$ is a clean speech signal and $n(t)$ is the background noise. To use the ICA technique, this method first segments the received signal $x(t)$ with time-domain window and generate the segments as the columns of matrices, to be exact,

$$X = S + N \tag{10}$$

where, the matrices X, S and N are of size $M \times K$, M is the speech frame size in samples from the pre-processing framing method and K is the number of frames. The remaining of the obtained signal, specifically smaller than a segment size is not taken into to account during the process of reshaping.

The speech has particular higher order statistical features [19]. With no loss of generality, this work possibly will presume that clean speech signal is completely the linear mixture of some independent constituents. ICA completely converts a collection of observed speech segments $sg = [sg_1, sg_2,\ldots, sg_M]^T$, that forms a column of matrix S into a new representation $SG = [SG_1, SG_2,\ldots, SG_M]^T$, where the components SG_i, $1 \le i \le T$ of SG are jointly statistically independent, that is,

$$SG = W.s \tag{11}$$

With W, an $M \times M$ invertible matrix is called unmixing matrix. By applying W from left side to a column of every matrix in Eq.(10), this will provide the following,

$$\beta = Wx = Ws + Wn = SG + \vartheta \tag{12}$$

where, x and n stands for one of the columns of matrix X and N, respectively, β and ϑ are the distorted speech segment and noise segment in ICA domain corresponding to x and n. All the above variables x, n, β and ϑ, are $M \times 1$ vectors. Let, SG denote the approximate SG in ICA domain based on x. By applying the inverse transformation to SG and obtained the enhanced speech segment vector, z, such as,

$$z = W^{-1}.SG. \tag{13}$$

These segmented enhanced speech vector is used for extract the signal features from the speech signal.

3.3 FEATURE EXTRACTION

The feature extraction module manipulate speech data so that further stages can use a more compact, though meaningful and tractable representation. Transforming the input data into a set of features is called feature extraction. During feature extraction step, a sequence of feature vectors are taken from original speech signal and unnecessary information from the signal are stripped and the properties of the signal which are important for the classification task, which are converted to a format that simplifies the distinction of the classes. The dimension of the data is reduced during feature extraction. So, feature extraction plays a vital role in speech processing. The feature vector sequences obtained are the inputs to the classification step of a speech processing system. The performance of speech processing in hearing aid application strongly depends on the accuracy of the feature extraction method. And this section proposes on Wavelet Transform based feature extraction method used for extracting features from speech signals. Feature extraction method based on Wavelet Transform most widely used method.

3.3.1 Discrete Wavelet Transformation Method for Speech signal feature extraction :

Wavelet transforms were introduced to address the problems associated with speech signal. A Wavelet transform decomposes a sound signal into a set of basic functions called wavelets. DWT is a special case of the wavelet transform that provides a compact representation of a speech signals in time and frequency that can be computed efficiently. They are well suitable for processing signals like speech because of their efficient time-frequency localization and the multi-resolutional, multi-scale analysis characteristics of the wavelet representations. The definition of DWT is described in following equation,

$$W(j,k)=\sum_{j}\sum_{k}x(k)2^{-\frac{1}{2}}\Psi\left(2^{-j}n-k\right) \tag{14}$$

where, $\Psi(t)$ is the basic time and energy analyzing function which is also called as mother wavelet.

The Wavelet Transform decomposes the speech signals over convert and expands mother wavelets. Mother wavelet is a time function with fixed energy and quick decay. In DWT, the frames of speech vector passes through two complementary filters, namely low-pass and high-pass filters, and emerges as two signals, called approximation coefficients and detail coefficients.

The low pass filter output is called as approximation coefficients and the highpass filter output is called as detail coefficients. In speech signals, the low frequency components characterize a signal more than its high frequency components and thus the low frequency components $h[n]$ are of greater importance than that of high frequency signals $g[n]$. DWT theory [20] needs following equations,

$$y_1[n] = \sum_{k=-\infty}^{\infty} x[k]h[n](2n-k) \text{ and} \tag{15}$$

$$y_h[n] = \sum_{k=-\infty}^{\infty} x[k]g[n](2n-k) \tag{16}$$

where, $h[n]$ represents an impulse response of a low-pass filter, and $g[n]$ indicates an impulse response of a high-pass filter. The scaling and wavelet functions can be applied efficiently by means of a pair of filters, i.e., $h[n]$ and $g[n]$ and these filters are regarded as a quadrature mirror filters that meet the property $g[n] = (-1)^{(1-n)}h[1-n]$ [21]. The input signal is low pass filtered in order to provide the fairly accurate features and high-pass filtered to provide the detail constituents of the input speech signal's segment vector. The fairly accurate speech signal vector during each stage is additionally decomposed by means of same low-pass and high-pass filters to obtain the approximate and detail characteristics for the subsequent stage such as classification in speech signal processing.

3.4 CLASSIFICATION OF THE EXTRACTED SPEECH SIGNAL VECTORS

In a classification, a classifier that takes an assessment based on the extracted features from the speech signal. In sound signal classification, the extracted feature vector as a given input to the classifier. Since the pure speech signal is computationally demanding to use directly, the relevant information is extracted from the speech signal.

The most basic classifier is the Bayesian, which is combined with ANFIS classifier, i.e. known as Modified classifier used in this proposed work for classifies the speech vector based on their extracted features. The extracted feature vector, $z = (z_1,\ldots z_k)$ and the classifier analyses the feature vector and takes a decision among N_d different decisions. The feature vector is processed through N_d different scalar-valued discriminant functions. The index, $d \in \{1,\ldots N_d\}$ to the discriminant function with the largest value given the observation is generated as output. A common special case is Maximum a Posteriori (MAP) decision where the classification task is to simply guess the source state with minimum error probability.

The source has a priori source state distribution $P(S = j)$ and this distribution is assumed to be known. The feature vector distribution is described by a conditional probability density function, $f_{(z|S)}(Z = z|S = j)$ and these conditional distributions are also assumed to be known. In practice, both the a priori distributions and the conditional distributions are normally unknown and must be estimated from training data. The probability that the source is in state $S = j$ given the observation $z = (z_1,\ldots z_k)$ can now be calculated by using Bayes' rule as:,

$$P_{(S|Z)}(j \mid z) = \frac{f_{(Z|S)}P_s(j)}{f_z(z)} \tag{17}$$

The decision function, the index of the discriminant function with the largest value given the observation, is:

$$d(z) = \arg\max_j g_j(z) \tag{18}$$

For a given observation sequence it was assumed here that the internal state in the source is fixed when generating the observed data sequence.

In many processes the input in the source will change describing a discrete state sequence; $S = (S)(1),\ldots,(S)(2),\ldots,(S)(T)$. In this case the observation sequence can have time-varying characteristics. Different states can have different probability density distributions for the output signal. Even though the Bayesian framework can be used when the input changes during the observation period, hence this Bayesian method combined with ANFIS which creates a modified classifier. A better modified classifier method, which includes a simple fuzzy neural network model of the dependencies between the input and the output. This modified model provides an efficient classification result for hearing aid application speech analysis process.

An ANFIS system could be used with fuzzy neural network model for fine adjustment of Bayesian classifier. Such framework makes the ANFIS modelling more efficient and less dependent on proficient knowledge and consequently assists learning and modification. The structure of ANFIS is shown in Fig.2. In the primary layer, the entire nodes are adaptive nodes. The outputs of layer 1 indicate the fuzzy membership grade of the inputs, which are given as [22]:

$$O_i^1 = \mu_{Ai}(x), O_i^1 = \mu_{B(i-1)}(x) \tag{19}$$

where, x stands for the input to node i and A_i and B_{i-1} stands for the linguistic label related with this function.

O_i^1 indicates the i^{th} output of layer 1, $\mu_{Ai}(x)$ and $\mu_{B(i-2)}(x)$ are type A and type B arbitrary fuzzy membership functions of nodes i and $i-2$, correspondingly. In case of the second and third layer, the nodes are predetermined nodes. They are labelled as M and N respectively, representing they carry out as a simple multiplier. The outputs of these layers can be indicated as,

$$O_i^2 = w_i = \mu_{Ai}(x)\mu_{Bi}(x) \tag{20}$$

$$O_i^3 = \overline{w_1} = \frac{w_i}{w_i + w_{i+1}} \tag{21}$$

where, $\overline{w_1}$ stands for the normalized firing strengths. In case of the fourth layer, the nodes are extremely adaptive nodes. The output of every node in this layer is basically the product of the normalized firing strength and a first order polynomial for the first order Sugeno model. The outputs of this layer are given as follows:

$$O_i^4 = \overline{w_1} f_i = \overline{w_1}(p_i x + q_i y + r_i). \tag{22}$$

where, f_i stands for the firing rate, p_i stands for the x scale, q_i stands for the y scale, and r_i indicates the bias for i^{th} node. In case of the fifth layer, there is only one single permanent node that carries out the summation of the entire incoming speech signals:

$$O_i^5 = \sum_{i=1}^{2} \overline{w_1} f_i = \sum_{i=1}^{2} \frac{w_i f_i}{w_i + w_{i+1}}. \tag{23}$$

Based on the proposed ANFIS architecture in Fig.2., it is found that provided the values of basis parameters, the overall output can be given as a linear groupings of the consequent outputσ, The process of the learning scheme for this structure is to fine-tune all the above mentioned modifiable parameters in order to make the ANFIS output equivalent to the training data.

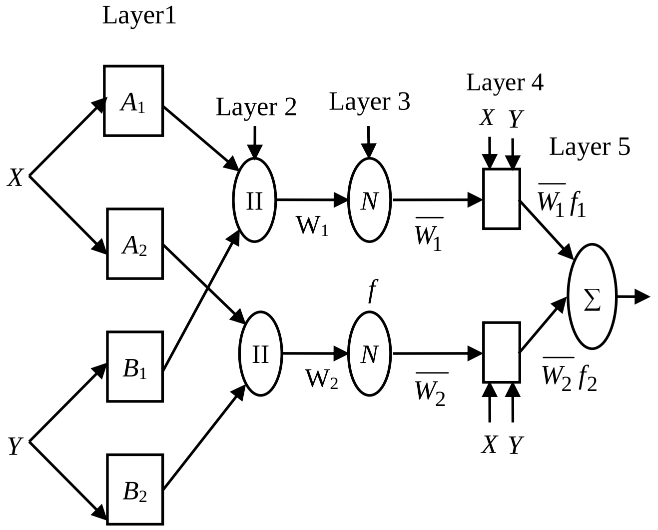

Fig.2. ANFIS architecture

In case if the premise parameters of the membership function are predetermined, the output of the ANFIS model can be given as follows:

$$\begin{aligned} \sigma &= \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \Rightarrow \overline{w_1} f_1 + \overline{w_2} f_2 \\ &\Rightarrow \left(\overline{w_1} x\right) p_1 + \left(\overline{w_1} y\right) q_1 + \left(\overline{w_1}\right) r_1 \\ &\quad + \left(\overline{w_2} x\right) p_2 + \left(\overline{w_2} y\right) q_2 + \left(\overline{w_2}\right) r_2 \end{aligned} \tag{24}$$

which is linear in the subsequent parameters (p_1, q_1, r_1, p_2, q_2 and r_2). In case if the premise parameters are predetermined the least squares scheme is employed without any difficulty to recognize the optimal values of these parameters subsequent to modification of ANFIS weights by means of Bayes functions. In case if the premise parameters are not predetermined, the search space happens to be larger and the convergence of the training becomes slower.

A speech signal classification scheme based on ANFIS integrated with Bayesian is proposed in this work. It is observed that feature vectors from the feature extraction process in the speech analysis process employed for the purpose of improving the classification rate as an effective component. The hearing aid application accomplished with the Bayesian time varying features is modified with the assistance of ANFIS. This work can be a strong foundation for improved schemes in the field of speech analysis, in case of the existence of the background noise for hearing aid applications.

4. RESULTS AND DISCUSSION

The performance analysis of this speech signal enhancement and classification method done with use of a clear speech signal, $s(t)$ = 1,2,…N and noise signal $n(t)$ = 1,2,.., N. This proposed method mainly focused on the performance improvement of hearing aid application where it is present in the number of background noise environments. In order to increase the performance of the hearing aid application, the background noises in the sound signal of the hearing aid application are must be reduced. Here this proposed speech enhancement and classification processes reduced the background noises from the noisy speech signal and classified the clear speech signal in efficient manner. The initial process of this proposed methodology used pre-processing and ICA noise reduction methods, which are applied to enhance the noisy speech signal in the hearing aid application and after that to extract the speech signal features for efficient classification process, here this methodology used DWT for feature extraction and finally Modified Bayesian ANFIS method is used to classifies the noise and clear speech signals in the hearing aid application. The classified clear speech signal proves that this proposed approach efficiently increase the speech intelligibility and quality in the hearing aid application in order to classifies the clear speech signal from different background noise environment. In case of the clear speech classification, the overall measure is obtained Segmental Signal-to-Noise Ratio, Perceptual Evaluation of Speech Quality (PESQ), Weighted Spectral Slope (WSS) and Articulation Index (AI) performance parameters.

In order measure the performance of the proposed methodology with use of above mentioned parameters, the RWCP- sound scene database (SSD) is used, which is a database of sounds in real acoustical environments. The performance parameter Segmental Signal to Noise ratio (SNR_{Seg}) is required for better quality enhanced speech signal and these measures are an objective quality measures corresponding to each frame of speech signal and Perceptual Evaluation of speech quality (PESQ) is subjective quality measure parameter; this is estimated with the help of various listening tests and Mean opinion score (MOS) of corresponding tests. PESQ measures suitable mainly for predicting signal distortion, noise distortion and overall speech quality. In the proposed methodology the clean and noise signals were processed by the ICA noise reduction algorithm, DWT feature extraction and modified classifier which is to simulate the receiving frequency characteristics of hearing aid application in the optimal way.

The performance evaluation of the speech enhancement and classification are compared with existing algorithms like Spectral subtraction [23], genetic SVD [24], and Deep Neural Networks (DNNs) [25], Empirical Mode Decomposition (EMD) with Adaptive Centre Weighted Average (ACWA) filter [26] and proposed Modified Bayesian-ANFIS which are cover the major classes of noise reduction in the speech analysis process. The proposed methodology is simulated on MATLAB, the simulation results of the proposed and existing method which shows that the efficiency of the proposed method's in clear speech classification. The detail description of the performance results of this work is explained as follows.

4.1 SEGMENTAL SIGNAL TO NOISE RATIO (SNR_{seg})

Segmental Signal-to-Noise Ratio, instead of working on the whole signal, calculates the average of the SNR values of short segments is given by,

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\sum_{i=Nm}^{Nm+N-1}\left(\frac{\sum_{i=1}^{N}x^2(i)}{\sum_{i=1}^{N}\left(x(i)-y(i)\right)^2}\right) \quad (25)$$

where, N is the segment length or frame length and M is the number of segments (frames) in the signal respectively. The length of segments is generally used in the performance parameter of Segmental SNR, which is discussed in [27]. The Segmental Signal-to-Noise Ratio is the most widely used for estimate the speech quality; hence here this work used this performance parameter i.e., Segmental Signal-to-Noise Ratio to measure the classified speech signal quality in hearing aid application. In this work there are three types of noise were used to generate noisy speech signals with the different segmental SNR level such as 0dB, 5dB, 10dB and 15dB. The noises used for experiments are white, pink and babble noises and the parameters for computing are used segment length or frame length is 10-20ms. The result of these parameter value used in the segmental SNR provides best results.

The Fig.3 and Fig.4 show that the segmental SNR results across different speech processing algorithms under three background noises such as white, pink and babble and Segmental SNR level.

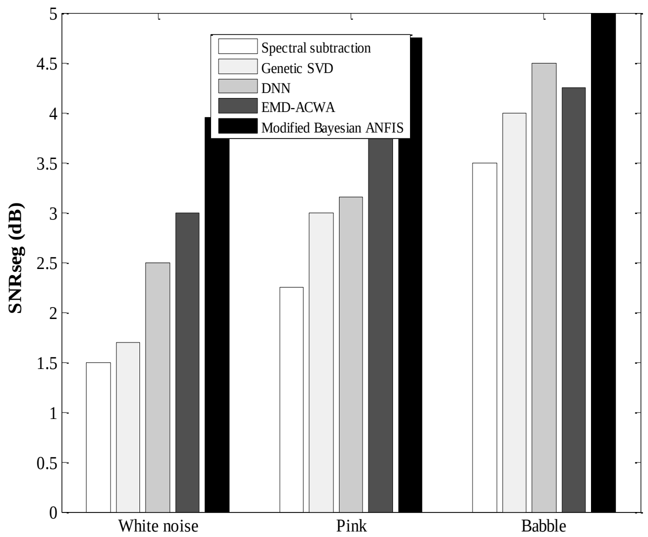

Fig.3. SNR_{seg} comparison results of the proposed and existing method present in the three different background noises

The proposed methods shows highest performance value for additive white, pink and babble background noise conditions. The Fig.3 and Fig.4 clearly shows the proposed Modified Bayesian ANFIS method shows the best Segmental SNR improvements than the existing methods in different SNR level such as 0dB, 5dB, 10dB, 15dB and segmental length value at 10-20ms.

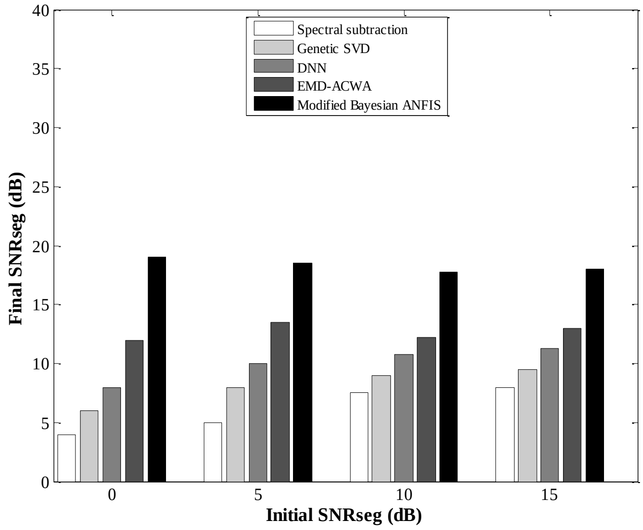

Fig.4. The comparison results of the final SNR_{seg}

4.2 PERCEPTUAL EVALUATION OF SPEECH QUALITY (PESQ)

The PESQ measure is the most complex to compute, and it is recommended by ITU-T for speech quality assessment. The final PESQ score is obtained by a linear combination of the average disturbance value D and the average asymmetrical disturbance values as follows,

$$PESQ = a_0 - a_1.D - a_2.A \quad (26)$$

where, $a_0 = 0.1$, $a_1 = 0.1$ and $a_2 = 0.0309$.

The Fig.5 and Fig.6 shows that the PESQ results across different speech processing algorithms under three background noises and PESQ levels. The Fig.5 and Fig.6 clearly shows the proposed method provides better results compared with existing methods when using $a_0 = 0.1$, $a_1 = 0.1$ and $a_2 = 0.0309$ values to estimate the speech quality of the hearing aid application.

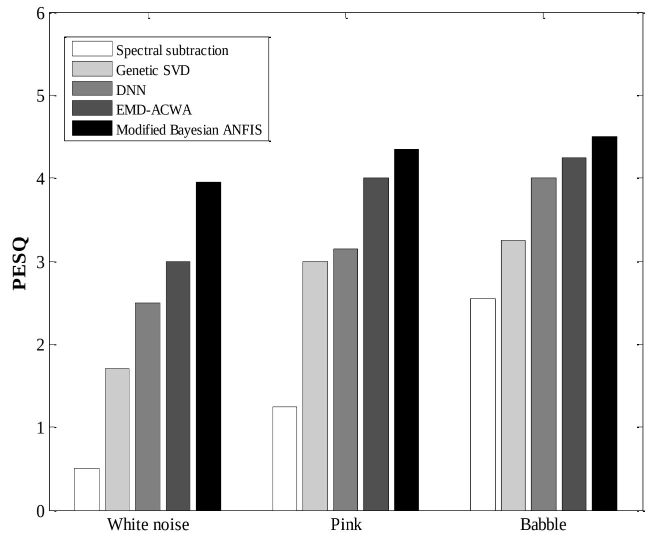

Fig.5. PESQ comparison results of the proposed and existing method present in the three different background noises

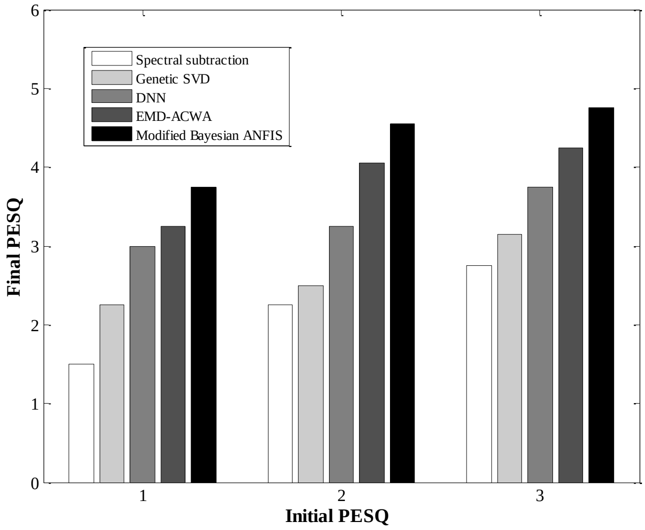

Fig.6. The comparison results of the final PESQ

4.3 WEIGHTED SPECTRAL SLOPE (WSS)

The Distance measure Weighted Spectral Slope (WSS) is a direct spectral distance measure. Based on the comparison of smoothed spectra from the clean and distorted speech samples the WSS is calculated, and it is defined as,

$$WSS = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} w_i\left(S_c - S_p\right)}{\sum_{j=1}^{K} w_i} \quad (27)$$

where, w_i are the weights calculated in [28], M is the number of speech signal segments, K is the number of bands, and S_c and S_p are the spectral slopes of the j^{th} frequency band in the m^{th} frame for noise-free and distorted speech signals respectively, which is typically known as the spectral differences between neighbouring bands. The Fig.7 shows that the WSS performance efficiency of the proposed which is high than the existing methods.

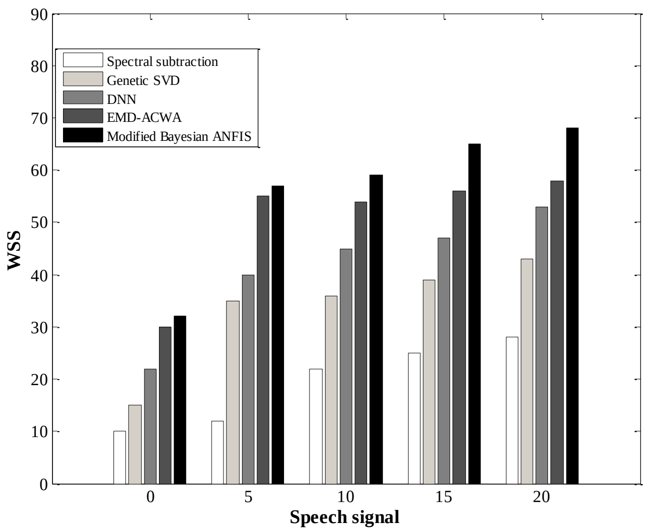

Fig.7. Performance of the WSS against the speech signal using proposed and existing methods

4.4 ARTICULATION INDEX (AI)

One more commonly accepted quality measure parameter is Articulation Index (AI), which used in this proposed speech enhancement and speech classification method that can estimate speech intelligibility of this more proposed method in optimal way. AI assumes that distortion signal can be calculated on a per-critical frequency band basis, and distortion in one frequency band does not affect other bands. The distortion is assumed to be either additive noise, or signal attenuation. AI can be obtained by calculating the SNR for each band, and averaging them as follows,

$$AI = \frac{1}{n}\sum_{j=1}^{n}\frac{\min\left\{SNR\left(j\right), m\right\}}{m} \quad (28)$$

where, $SNR(j)$ is the SNR of the j^{th} subband, n the is number of subbands, and m is denoted as the maximum subband SNR. In this work set to be n =10 and the maximum subband SNR, m is set to 20dB, which provide to be the best result value as shown in Fig.8. The contribution of each band is set to uniform in this case and the maximum subband SNR can be set to different values, and different weights for each band can be set as well for the calculation of speech intelligibility used in this proposed wok.

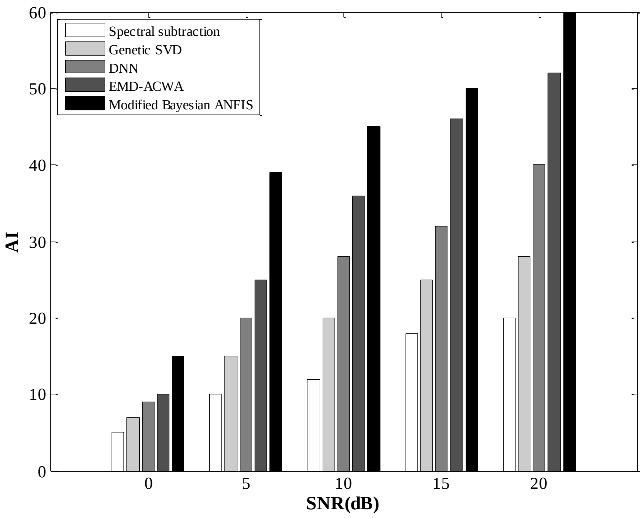

Fig.8. Relationship between the AI and speech intelligibility comparison result against proposed and existing methods

The Fig.8 clearly shows that the speech intelligibility parameter AI of the proposed work is performed well than the existing method. It will shows that the proposed classification method efficiently classified the clean speech signal from the noisy environment in the hearing aid application.

The above shown intelligibility and speech quality measurement results reflect the true performance of proposed speech enhancement algorithm and classification methods used in this work. The Speech Quality assessment is done using different quality measures and these are compared with existing techniques and the result graphs shows that the proposed Modified Bayesian ANFIS method with use of ICA noise reduction technique and pre-processing and DWT feature extraction techniques performed moderately well on classifying speech signal from different background noises present in hearing aid applications and the results shows that the proposed method has a better performance compare than the existing methods.

5. CONCLUSION

In speech signal analysis of hearing aid applications, the classification process is most important step to classify the background noises and enhancing the speech signals. To classify number of background noises present at a situations in hearing aid application, the speech signals are correctly classified with the use of proposed Modified Bayesian with ANFIS's, which can take advantage of the result requires information about the sound frequency direction, and the number, distance, and type of sound sources in the room or outside. This information can be derived from classification algorithm. The experimentation results show that this proposed method using Independent Component Analysis (ICA) for reducing the background noise and the Discrete Wavelet Transforms could effectively extract the features from the speech signal and Modified Bayesian-ANFIS classify it for efficient manner to increase the speech quality in the hearing aids. Future work will involve the modification of the proposed method, such as employing the optimal number of states and components for different background noises, so that the evaluation is more robust and further identifying different kinds of environment sounds belonging to the "others" sound class.

REFERENCES

[1] Baker J. Richard and S. Rosen, "Auditory Filter Nonlinearity in Mild/Moderate Hearing Impairment", *The Journal of the Acoustical Society of America*, Vol. 111, No. 3, pp. 1330-1339, 2002.

[2] Gerasimos Potamianos, C. Neti, G. Gravier, A. Garg and A.W. Senior, "Recent Advances in the Automatic Recognition of Audio Visual Speech", *Proceedings of the IEEE* , Vol. 91, No. 9, pp. 1306-1326, 2003.

[3] Hagai Attias, John C. Platt, Alex Acero and Li Deng, "Speech Denoising and Dereverberation using Probabilistic Models", *Proceedings of Advances in Neural Information Processing Systems*, Vol. 13, pp. 758-764, 2001.

[4] Israel Cohen, Sharon Gannot and Baruch Berdugo, "An Integrated Real-Time Beamforming and Postfiltering System for Nonstationary Noise Environments", *Journal on Applied Signal Processing*, Vol. 2003, pp.1064-1073, 2003.

[5] Kuldip Paliwal, Belinda Schwerin and Kamil Wójcicki, "Speech Enhancement using a Minimum Mean-Square Error Short-Time Spectral Modulation Magnitude Estimator", *Speech Communication*, Vol. 54, No. 2, pp. 282-305, 2012.

[6] Y. Ephraim, "Statistical-Model-Based Speech Enhancement Systems", *Proceedings of the IEEE*, Vol. 80, No. 10, pp. 1526-1555, 1992.

[7] Trausti Kristjansson and John Hershey, "High Resolution Signal Reconstruction", *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 291-296, 2003.

[8] Alberto Abad Gareta, "A Multi-Microphone Approach to Speech Processing in a Smart-Room Environment", Ph.D Thesis, Department of Signal Theory and Communications, Polytechnic University of Catalonia, 2007.

[9] Pierre Comon, "Independent Component Analysis, a New Concept?", *Signal Processing*, Vol. 36, No. 3, pp. 287-314, 1994.

[10] Y. Huang, J. Benesty and J. Chen, "Analysis and Comparison of Multichannel Noise Reduction methods in a Common Framework", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 5, pp. 957-968, 2008.

[11] Michael Brandstein and D Ward, "*Microphone Arrays: Signal Processing Techniques and Applications*", Springer, 2001.

[12] Hadi Veisi and Hossein Sameti, "Speech Enhancement using Hidden Markov Models in Mel-Frequency Domain", *Speech Communication*, Vol. 55, No. 2, pp. 205-220, 2013.

[13] T. Van Den Bogaert, S Doclo, J Wouters and M Moonen, "Speech Enhancement with Multichannel Wiener Filter Techniques in Multimicrophone Binaural Hearing Aids", *The Journal of the Acoustical Society of America*, Vol. 125, No. 1, pp. 360-371, 2009.

[14] Y.W. Lee, S.M. Lee, Y.S. Ji, J.S. Lee, Y.J. Chee, S.H. Hong and I.Y. Kim, "An Efficient Speech Enhancement Algorithm for Digital Hearing Aids based on Modified Spectral Subtraction and Companding", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E90-A, No. 8, pp. 1628-1635, 2007.

[15] N.A. Shusina and B. Rafaely, "Unbiased Adaptive Feedback Cancellation in Hearing Aids by Closed-Loop Identification", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 2, pp. 658-665, 2006.

[16] H. Li and X. Hong, "Binaural Auditory Localization of Signals Processed by speech Enhancement Methods", *Proceedings of 7th International Congress on Image and Signal Processing*, pp. 883-887, 2014.

[17] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, Vol. 13, No. 4-5, pp. 411-430, 2000.

[18] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao and F. Yang, "BCI Competition 2003-Data Set IIb: Enhancing P 300 Wave Detection using ICA-based Subspace Projections for BCI Applications", *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 6, pp. 1067-1072, 2004.

[19] Jong-Hwan Lee, Ho-Young Jung, Te-Won Lee and Soo-Young Lee, "Speech Feature Extraction using Independent Component Analysis" , *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. 1631-1634, 2000.

[20] Yu Hao and Xiaoyan Zhu, "A New Feature in Speech Recognition based on Wavelet Transform", *Proceedings of 5th International Conference on Signal Processing*, Vol. 3, pp. 1526-1529, 2000.

[21] Stephane Mallat, "*A Wavelet Tour of Signal Processing*", Academic press, 1999.

[22] J.S. Roger, "ANFIS: Adaptive-Network-based Fuzzy Inference System", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 23, No. 3, pp. 665-685, 1993.

[23] E. Verteletskaya and B. Simak, "Noise Reduction based on Modified Spectral Subtraction Method", *International*

Journal of Computer Science, Vol. 38, No. 1, pp. 82-88, 2011.

[24] A. Zehtabian, H. Hassanpour, S. Zehtabian and V. Zarzoso, "A Novel Speech Enhancement Approach based on Singular Value Decomposition and Genetic Algorithm", *Proceedings of International Conference of Soft Computing and Pattern Recognition*, pp. 430-435, 2010.

[25] Y. Jiang, D. Wang, R. Liu and Z. Feng, "Binaural Classification for Reverberant Speech Segregation using Deep Neural Networks", *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* Vol. 22, No. 12, pp. 2112-2121, 2014.

[26] K. Khaldi, A.O. Boudraa and M. Turki, "Voiced/Unvoiced Speech Classification-Based Adaptive Filtering of Decomposed Empirical Modes for Speech Enhancement", *IET Signal Processing*, Vol. 10, No. 1, pp. 69-80, 2016.

[27] Wonho Yang, "Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure based on Audible Distortion and Cognition Model", Ph.D Dissertation, Temple University, 1999.

[28] D. Klatt, "Prediction of Perceived Phonetic Distances from Critical Band Spectra", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 7, pp. 1278-1281, 1982.