IJCTA, 9(2) 2016, pp. 1367-1380

Speech Signal Enhancement Using Fast Ica With Optimised Fa-anfis Classifier for Hearing AIDS Application

N. Shanmugapriya[1] and E. Chandra[2]

ABSTRACT

In hearing aids application a long-standing challenge is to improve the understanding of speech in presence of background noises. Hearing aids are well supported the speech enhancement methods, which are actually help the end listener to understand the speech in various background noisy environments. These speech enhancement methods also increase the speech intelligibility and speech quality in the hearing aids application. Here this work proposed an approaches to improve the speech intelligibility and quality in the hearing aids based on the following enhancement methods, such as Fast Independent Component Analysis (Fast ICA) algorithm, which is used for noise reduction from the noisy speech signal, then this work used Improved Discrete Wavelet Transform (IDWT) for feature extraction from denoised speech signal and Advanced Analysis of Variance (AANOVA) method used for select the significant feature from extracted features of the speech signal then this selected features are optimized with Firefly algorithm (FA) and then the optimized features are efficiently classify with use of Adaptive Neural Fuzzy Interference System (ANFIS) classification method, i.e. Optimized FA-ANFIS Method. The experimental result of the proposed method has to prove and it provide a better speech intelligibility and high quality of the speech signal for Hearing aids application than compared to the existing methods.

Key Words: Speech enhancement, Fast Independent Component Analysis (Fast ICA) algorithm, Improved Discrete Wavelet Transform (IDWT), Optimized FA-ANFIS Method, speech intelligibility and high quality.

1. INTRODUCTION

Speech enhancement involves processing speech signals for human listening or as preparation for further processing before listening. In the hearing aids application the main goal of enhancement process is to improve the overall quality of speeches; and classify the noise from clean speeches to increase the speech clearness in order to reduce the hearing aids user's exhaustion, ambiguity etc, depending on specific application. The enhancement system may be designed to achieve one of these aims or several. Perhaps the most common complaint of hearing aids users is that of background noise. In hearing aids application for speech signal enhancement, there is a need to separate the mixture of speech signals or extract a clear speech signal of interest while reducing undesired interfering speech signals and background noises. The estimated signals may then be either directly listened to or further processed, giving rise to a wide range of other different applications such as human computer interaction, surveillance, and hands-free telephony[1]. Persons with normal hearing can use binaural clues to help focus on the desired sounds while attenuating noise, but even this ability is sometimes insufficient. For the hearing impaired people, the problem increases significantly. To overcome this problem in hearing aids application to enhance the speech signal with the background noise reduction. The speech enhancement method is also increased intelligibility and quality of the speech signal in hearing aid application.

[1] Assistant Professor, Department of Computer Applications (MCA), Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore -641 049, Tamil Nadu, India, *E-mail: spriyanatrajan@gmail.com*

[2] Professor and Head, Department of Computer Science, Bharathiar University, Coimbatore-641 046, Tamil Nadu, India, *E-mail: crcspeech@gmail.com*

For speech enhancement, there are available different Digital signal processing (DSP) techniques, which include spectral subtraction [2]–[4], adaptive filtering [5], [6], [7] and suppression of non-harmonic frequencies [6]–[8]. Most of these techniques either require a second microphone to provide the noise reference [5], [9], [10], or require that the characteristics of noise be relatively stationary. Clearly the speech signal in hearing aids application is non-stationary. The proposed speech enhancement method with noise reduction, feature extraction and feature selection and classification method to handle the non-stationary speech signal more efficiently with different no. of background noises. The basic advantage of this method is the implementation simplicity and relatively light computation requirements. Hence this work have developed a classification based with signal enhancement system using signal processing method for increase the speech intelligibility and quality of the speech in hearing aid applications, which will be described in detail about the following sections.

The proposed methods of this work organized as follows, a related work of the proposed speech enhancement process is described in Section 2, In Section 3, proposed methodology of the speech enhancement systems is described, there are five methods developed in this work aiming at improving the performance efficiency of the hearing aids application with optimized FA-ANFIS based enhancement system are presented. In Section 4. The experimentation results of the proposed method comparing various types of the speech enhancement systems. Finally, Section 5 encloses the conclusion of this proposed work.

2. RELATED WORK

In [11] proposed a speech enhancement process, which is concerned with the processing of corrupted or noisy speech signal in order to improve the quality or intelligibility of the signal. The main aim of this proposed work is to enhance speech signal corrupted by noise to obtain a clean signal with higher quality. However, the presence of noise in speech signals will contribute to a high degree of inaccuracy in a system that requires speech processing. The noise cancellation process for the speech signal was preceded with the neural networks. There are three methods were used to test, first one is the Adaptive Linear Neuron (ADALINE), then Feed Forward Neural Network Enhancement Method (FFNN), finally, Wavelet Transform and Adaline Enhancement Method. The results from this proposed work shows the high quality due to fast processing and high signal-noise-ratio.

In [12] presented the application of an Adaptive Wiener in speech enhancement. The proposed filter depends on the adaptation of the filter transfer function from sample to sample based on the speech signal statistics (mean and variance). The proposed Wiener filter is implemented in time domain rather than in frequency domain to accommodate for the varying nature of the speech signal. The proposed method is compared to the traditional Wiener filter and the spectral subtraction methods and the results reveal its advantage of the proposed adaptive wiener filter.

In speech processing, speech signal enhancement is an important topic. There is some problem, which affects the signal enhancement process. This problem reduces the quality of speech and audio signals. In [13] discussed and experimented about the some adaptive noise cancellation algorithms, which are used to reduce the noise with relatively fast convergence as preferred for speech processing, the experimented noise reduction techniques are Least Mean Square (LMS), Normalized LMS (NLMS) and Recursive Least Squares (RLS) are widely used easy computation and implementation methods. These algorithms are evaluated with following conditions, sensitivity for language, text gender and noise power. From analysis [13] concluded that the RLS provides the best performance under noise power is fixed and that noise power has more influence on this algorithm.

In [14] examined the speech signal enhancement using Kernel Affine Projection Algorithm (KAPA) and Normalized KAPA. In order to improve the quality and intelligibility of speech, unlike time and frequency domains, in this proposed work process the signal in new domain like Reproducing Kernel Hilbert Space

(RKHS) for high dimensional to yield more powerful nonlinear extensions. This proposed experimental work used the database of noisy speech corpus (NOIZEUS). The experimental result showed the RKHS is provides the better noise removal result than the existing methods.

In [15] presented the Bayesian paradigm, which provides a natural and effective means of exploiting prior knowledge concerning the time-frequency structure of sound signals such as speech and music something which has often been overlooked in traditional audio signal processing approaches. Here, after constructing a Bayesian model and prior distributions capable of taking into account the time-frequency characteristics of typical audio waveforms applied Markov chain Monte Carlo methods in order to sample from the resultant posterior distribution of interest. This work presented a speech enhancement results which compare favorably in objective terms with standard time-varying filtering techniques (and in several cases yield superior performance, both objectively and subjectively); moreover, in contrast to such methods, the proposed method's result is obtained without an assumption of prior knowledge of the noise power.

3. PROPOSED METHODOLOGY

In hearing aids applications, speech intelligibility and quality is most challenging task when there are number of back ground noise environment present in the speech signal. To improve the hearing aids application performance and speech quality, the proposed speech signal processing methodology in hearing aid application is depicts as follows:

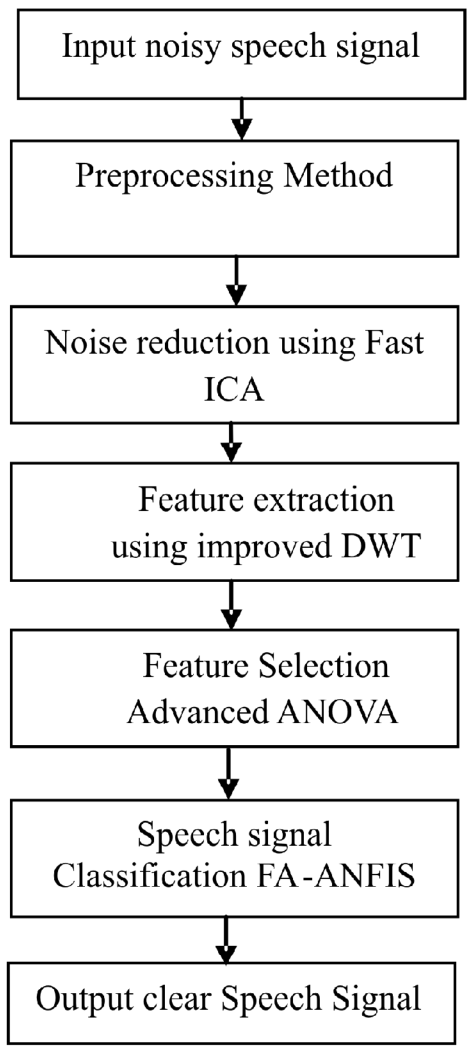

Figure 1: The flow diagram for Speech signal Process

3.1. Pre-processing method

A speech pre-processing method is presented to improve the speech intelligibility in noise environment for the near-end listener in hearing aids applications for hearing impaired users. In this section described in detail about the process of preprocessing technique for a non-stationary speech signals with background noises.

A speech signal is denote as $s(t)$ and $n(t)$ is a background noise and noisy speech signal represents as $x(t)$. The distortion measure considered in this proposed work, denoted by $d(x\ (t))$, shows the audibility of $n(t)$ in the presence of $s(t)$. Hence, a lower $d(x(t))$ value implies less audible noise and therefore most clear audible speech in hearing aids application. The main aim of this preprocessing step is to adjust the speech signal $s(t)$ such that d is minimized subject and modified clear speech remains unchanged.

Proposed Pre-processing method for speech signals

The perceptual distortion measure is based on the work from [16], which takes into account a non stationary speech signals for processing. Before this distortion measure for non-stationary speech signals for speech intelligibility, framing and windowing process is applied to the noisy speech signal $x(t)$. For distortion measure process considers the sequential wrapping within a short-time frame from the framing and windowing method. Consequently, for speech intelligibility the distortion measure is evaluated for redistributing the speech energy over time and frequency, distortion measure is more importance and it is also sensitive to transients.

Framing

In speech processing it is often advantageous to divide the signal into frames to achieve stationary, because the speech signal is generally in non- stationary form. In this framing step the speech signal $s(t)$ 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 by $X – Y$ samples. This process continues until all the 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 signal i.e. background noises of the hearing aids application.

Windowing

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

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

A typical window used in the speech signal enhancement is Hamming window, which has the form given by equation (2).

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

After windowing and farming methods are applied to the noisy speech signal, that speech signals are now used in distortion measure evaluation process, which is used a Time frequency unit of the preprocessed signal with less speech energy usage of the proposed system, which will help to increase the speech clearness of the hearing aids application while it is used in noisy environment.

Measurement of perceptual distortion for preprocessing

Let consider y_p is the impulse response of the p^{th} hearing filter and x_m the m^{th} short-time frame of the noisy speech signal $x(t)$, their linear convolution is denoted by $x_{p,m} = x_m * y_p$. Subsequently, the temporal envelope is defined by $|x_{m,p}|^2 * y_s$, where y_s represents the smoothing low-pass filter. The cutoff frequency of the low-pass filter determines the sensitivity of the model towards temporal fluctuations within a short-time frame. The audibility of the noise in presence of the speech signals, within one Time Frequency-unit, is determined by a per-sample noise-to-signal ratio [16]. An intermediate distortion measure for one TF-unit is obtained denoted by lower-case d. The distortion measure for the complete speech signal is then obtained by summing all the individual distortion outcomes over time and frequency is determined as follows,

$$d\left(x_{m,p}\right) = \sum_n \left|x_{p,m}\right|^2(n) = d(x_m * y_p), \tag{3}$$

Power-Constrained Speech-Audibility Optimization in Hearing aids application

To improve the speech audibility when presence of noise, use to minimize Equation (3) by applying a gain function α which redistributes the energy of the speech signal, i.e., $\alpha_{m,p}$, $x_{m,p}$, where $\alpha_{m,p} \geq 0$. Only TF-units are modified where speech is present. This is done in order to prevent that a large amount of energy would be redistributed to speech-absent regions.

This work considers a TF-unit to be speech-active in hearing aids application, when its energy is particular range of the TF-unit with maximum energy within that particular frequency band of the signal. The background noise of the speech signal $x(t)$ is assumed to be a stochastic process, which is denoted by $n_{m,p}$ and the speech deterministic. Hence minimize for the expected value of the distortion measure. In order to determine a gain function α, which have to evaluate the expected value $E[d\ (x_{m,p})]$ which can be expressed as follows,

$$E\left[d\left(x_{m,p}\right)\right] = \sum_n \left|x_{p,m}\right|^2(n) \tag{4}$$

To simplify assume that the power-spectral density of the noise within the frequency range of an acoustic band is constant, i.e., has a ’flat’ spectrum. As a consequence, the noise within an acoustic band can be modeled by $n_{m,p} = (w_m, N_{m,p}) * y_p$, where w_m denotes the window function and $N_{m,p}$ represents a zero mean, i.e. stochastic process with variance $E\,(N^2{}_{m,p}\,(n) = \sigma^2_{m,p}, \forall n$. Here $\sigma^2_{m,p}$ is estimated with the noise power-spectral density (PSD) estimator from [17] by taking the average PSD within an acoustic band. As a final step of this preprocessing, an exponential smoother is applied to $\alpha_{m,p}$ in order to prevent ’background noise’ which may negatively effect the speech signal quality.

$$\hat{}\alpha_{m,p} = (1 - \gamma)\alpha_{m,i} + \gamma\ \hat{}\alpha_{m-1,p} \tag{5}$$

where $\gamma = 0.9$. To reduce complexity, the filter bank and the low-pass filter are applied by means of a point-wise multiplication in the DFT-domain with real-valued, even-symmetric frequency responses.

Thus the speech pre-processing method is presented in this work to improve the speech intelligibility in noisy environment. This was accomplished by optimally redistributing the speech energy over time and frequency based on a perceptual distortion measure.

After the pre-processing, speech signal $s(t)$ also contain the background noises, such as $n(t)$. The following enhancement processes of the speech signal processing such as noise reduction process is used to reduce the background noises present in the speech signal $s(t)$, then segment these speech signal based on the noises present in this signal, used this segmented speech signal, extract the features based spectral-temporal values, then select the significant feature for speech signal classification. Finally classification method classifies the speech signal, and this classified signal is more efficiently and audibly reaches to the end listener. Thus this process will increase the quality of the hearing aids application.

3.2. Proposed Noise Reduction Process for Hearing aids application

Signal processing methods in the hearing aids application allow for reducing the effects of noise. The recent development of digital hearing aids opens up substantial new possibilities with respect to the use of advanced signal processing techniques for noise reduction [18]. Because of the particularly damaging effects of background noise on speech intelligibility for people with hearing loss (i.e., hearing-aid users) this problem is of critical importance. On the positive side, background noise reduction process improves speech intelligibility in hearing aids application, that allow for the development of signal-processing strategies that may be of benefit to the hearing-aid user. The objective in applying noise techniques is not so much to reduce background noise, but to reduce the effects of background noise on speech intelligibility and overall sound quality [19]. In recent times, one of the most promising approaches to the problem of noise reduction is application of ICA. This novel ICA algorithm is robust with respect to additive noise, which makes it possible to use them successfully for noisy speech signal i.e. .

Independent Component Analysis

Before the ICA process analysis, the speech signal must be segmented. The segmentation is recognizing by a forced alignment of the speech with the other signal. Assume that speech enhance processing there exist segmented speech signals, $s_1(t)$, ..., $s_m(t)$ in Independent Component Analysis, that are scalar-valued and mutually (spatially) statistically independent, i.e. known as independent components (IC's) at each time instant or index value t. The clear speech signal sources $s_j(t)$ are unknown to the end listener when present the there is some noises, who also has to deal with possibly noisy signal but different linear noisy speech mixtures, $x_1(t)$,..., $x_n(t)$, of the sources usually for n greater than m. The mixing coefficients of the noisy signals $x(t)$ are some unknown constants.

Process of ICA

ICA process is to find the waveforms $s_i(t)$ of the clear speech signal sources, knowing only the noisy speech signal mixtures $x_j(t)$ and the number m of sources.

The noisy speech signal is denote by x(t) = [x1(t), ..., xn(t)]^T the n-dimensional t^{th} signal feature vector made up of the mixtures at discrete index value (usually time) t. The ICA mixing model is equal to:

$$x(t) = Ps(t) + n(t) = \sum_{i=1}^{m} s_i(t)p_i + n(t). \tag{6}$$

In general case the noise signal has a Gaussian distribution in assumption but none of the speech signals is Gaussian. To simplify this at most one of the Speech signals $s_i(t)$ is allowed to have a Gaussian distribution. These assumptions follow from the fact that it is impossible to differentiate several Gaussian and speech signals from each other. In typical noise reduction approach, a u × v separating matrix $w(t)$ is updated so that the u – vector, $z(t) = w(t)\ x(t)$, becomes an estimate of the original independent speech signals. $z(t)$ is the output feature vector of the original speech signal and the matrix $w(t)$ is the total weight matrix. In this proposed work used batch based approach fast ICA method for noise reduction process, which is describe as follows, before the fast ICA approach applied to the mixed noisy speech signal, there is preprocessing technique applied to reduce the no. of parameter estimated by Fast ICA.

Pre-processing in ICA

There is preprocessing method in ICA, known as Whitening, which is a linear transformation such that the observation feature vector elements will be uncorrelated and with unit variances:

$$E\{\tilde{x}\tilde{x}\} = I. \tag{7}$$

Whitening allows the reduction of the ICA search problem from n^2 free matrix coefficients to only $n(n–1)/2$ elements, as the matrix must be kept orthogonal. Even the matrix size reduction could be possible, if some eigen values λ_j, are too small. After whitening process, efficient fast ICA approach applied to the noisy speech mixture signal $x(t)$

Proposed Noise Reduction FastICA Process

An efficient batch (block) algorithm based FastICA method proposed by Hyvarinen et al. [20]. The noise reduction processing first allows a preprocessing whitening step for the original noisy speech mixture signals, which improves the processing speed of the ICA procedure. In this Fast ICA approach assume that the each coefficient of basis vectors have Gaussian distribution. Then the trained FastICA basis vectors can be applied to Maximum a posteriori (MAP) estimation. To reduce the noises from the noisy speech signal several steps are needed.

1. To estimate shrinkage function h_i of the i – th basis vector and randomize the initial weight vector w.
2. Let weight vector $w^+ \leftarrow E\{x_g(w^T x)^T\} - E\{g'(w^T x)\}w), where\ E\{\ldots.\}$ denotes the averaging over all column-vectors of matrix of noisy speech signal x, then calculate the noisy co-efficient vector z for the noisy input speech signal which is denoted as $\tilde{x}$, *whrere* $z = \tilde{w}x$. Let$w \leftarrow w^+/||w^+||$. Let, if not converged again averaging over all column-vectors of matrix of noisy speech signal.
3. Obtain the denoised coefficients from speech signal, $\hat{s_i} = h_i(z_i)$.
4. Finally, recover the denoised speech signal, $\hat{x} = w^{-1}\hat{s} = P\hat{s}$, where $\hat{s}$ the denoised basis vector coefficient, P is matrix.

After noise reduction process the spectro-temporal based features are extracted from this noise less speech signal $s(t)$. These extracted features are further used in the classification process in the speech signal analysis.

3.3. Feature extraction for speech signal process

The feature extraction method used in this proposed work for extract amount of features from the speech signal. In the feature extraction stage, from the preprocessed denoising speech signals are to extract the basic acoustic or linguistic features, such as pitch-related, intensity-related, duration-related, spectral-related or contour-related, tone-based and/ or vowel-related features. In addition, some transform functions are often employed to convert the speech features between different data domains, such time and frequency domains. The result of this stage is a speech data set represented by a set of high-dimensional speech features. The feature extraction method used in this work used wavelet based transform [21].

The analysis by the Wavelet Transforms is performed by using a single wave is called a mother wavelet (often called window) $\Psi(t)$ which can be considered as a band pass filter, has a limited duration of zero average, has irregular form, and often non-symmetrical unlike sinusoidal waves which extends from minus to plus infinity, and has a symmetrical form.

The wavelet transform is defined as the inner product of a signal $x(t)$ with the mother wavelet

$$\Psi(t)_{c\,d} = \Psi\,(\frac{t-c}{d}) \tag{8}$$

$$w_\Psi x(c,d) = \frac{1}{\sqrt{c}} \int_{-\infty}^{\infty} x(t)\Psi^*(\frac{t-c}{d}) \tag{9}$$

where c,d are the scale and translation parameters. Depending on the scale factor c, the Wavelet Transforms have a large freedom degree to vary window size. This varying window size is wide for slow frequencies because the low frequency component completes a cycle in a large time interval, and it is narrow for the other high frequencies as the high frequency component completes a cycle in a much shorter interval [22]. Thus, an optimal time-frequency resolution is obtained in all frequency ranges. The following section describes the proposed wavelet based improved features extraction method for speech analysis process.

Improved Discrete wavelet transform for feature extraction from speech signal

In general the scale and translation parameters of the DWT are denoted by, $c = c^j$, $d = kdc^j$, where k and j are integers. The function family of DWT becomes,

$$\Psi_{j,k}(t) = c^{-\frac{j}{2}}\Psi(c^{-j}t - kd) \tag{10}$$

Assumption that speech signals transformed by DWT as locating features into a series of high-pass and low-pass filters $g(n)$, $h(n)$ respectively. The high frequency content of speech signals which passed through the $g(n)$ filter is called "details", and the low frequency content which passed through the $h(n)$ filter is called "approximations", and only the approximation coefficients can be decomposed further as the decomposition level grows. The wavelet decomposition of the speech signal analyzed at level j.

The main drawback of the DWT is the high frequency components are removed while feature extraction, to overcome this problem in feature extraction process, this work proposed wavelet packet analysis is an improved method of the DWT [23]. Unlike the DWT, Improved DWT decomposes both the high and low frequency bands at each iteration. A pair of low pass and high pass filters is used to generate two sub-bands with different frequencies and this considered one level of decomposition. These sub-bands are then down-sampled dyadically. This process can be repeated to partition the frequency spectrum into smaller frequency bands for resolving different features while localizing the temporal information. This proposed improved DWT method used Daubechies wavelet for further analysis. The Daubechies wavelet is orthogonal wavelets. It has a maximum number of vanishing moments, and conserves the energy of the speech signal and redistributed in a more compact form. The energy corresponding to improved DWT method used each wavelet packet node is calculated by using the following equation according to the energy conservation of the speech signal, which is involved in the feature extraction process is as follows,

$$\sum_{t=-\infty}^{\infty} (|s(t)|)^2 \tag{11}$$

Thus the feature extraction process extracts the feature vector from speech signal with energy conservation, which is used for feature selection process.

3.4. Feature Selection Method for speech signal process

Feature selection process of this work is used to select relevant features in the temporal domain or relevant features in the spectral or cepstral domain, and simultaneously discard other noisy features speech signal in hearing aids application. The feature vectors from the feature extraction process may contain different amounts of spatio-temporal based information. A feature selection is necessary to obtain robust models for classification process. In this work proposed an Advanced Analysis of Variance (AANOVA) [24] for feature selection.

Advanced ANOVA

An advanced ANOVA is used to determine the significance of each feature in the speech signal use of multivariate feature vectors. It assumes that the observations are normally distributed and that the variances of the feature vectors are equal in all G groups. The means of a single feature vector for two or more groups are equal i.e. known as null assumption, which is rejected if the *p–value* as follows:

$$p - \text{value} = P(FG - 1, n - G > F - \text{ratio}) \tag{12}$$

is smaller than a chosen level of significance, α. *The F – ratio*

$$F - ratio = \frac{\sum_{G=1}^{G} n_G(\overline{f_G} - f_G)^2}{\sum_{G=1}^{G}(n_G - 1)\sigma_G^2} * \frac{n-G}{G-1} \tag{13}$$

can be thought of as a measure of how different the means are relative to the variability within each class, where $\bar{f_G}$ is the average of a single feature f within class G, f_G is the overall mean of the feature f within class G, σ_G^2 the variance and f the global mean. Fisher showed that, under the given assumptions, this ratio follows an F-distribution with $G - 1$ and $n - G$ degrees of freedom. The larger this value, the greater the likelihood that the differences between the domain specific means are due to something else than chance alone. So, the domain specific means of the feature with a small p-value are statistically unequal due to the accent variation instead of chance. The features with small p-values can be used as a feature set on which a following classifier is built to classify the speech signal based on the selected feature from the noisy signal and improve the hearing aids application quality.

3.5. Speech Signal Classification

In this work proposes an ANFIS model to classify the selected speech feature vector for speech signal enhancement. First, this work calculated the fitness values between the selected feature vectors of the speech signals. Then the ANFIS model is used to classify these feature vectors based on the fitness value of the selected speech signal features.

Firefly Algorithm

Firstly Xin-She Yang (2007 and 2008) was developed Firefly Algorithm (FA) [25], which was based on the flashing behaviors of fireflies.

In essence, FA uses the following three idealized rules:

- Fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.
- The attractiveness is proportional to the brightness, and they both decrease as their distance increases. Thus for any two flashing fireflies, the less brighter one will move towards the brighter one. If there is no brighter one than a particular firefly, it will move randomly.
- The brightness of a firefly is determined by the landscape of the objective function. As a firefly's attractiveness is proportional to the light intensity seen by adjacent fireflies, this can now define the variation of attractiveness β with the distance d by,

$$\beta = \beta_0 e^{-\gamma d^2} \tag{14}$$

where β_0 is the attractiveness at d = 0. The movement of a firefly x is attracted to another more attractive (brighter) firefly y is determined by,

$$a_x^{t+1} = a_x^t + \beta_0 e^{-\gamma d_{x,y}^2} \left(a_y^t - a_x^t\right) + \alpha_t \epsilon_x^t \tag{15}$$

where the second term is due to the attraction. The third term is randomization with being the randomization parameter, and it is a vector of random numbers drawn from a Gaussian distribution or uniform distribution at time t. If $\beta_0 = 0$, it becomes a simple random walk. Alternatively, if $\gamma = 0$, it reduces to a variant of particle swarm optimization [25].

ANFIS

The most basic ANFIS classifier combine with swarm intelligence based optimization algorithm is proposed for classifies the speech feature vectors for classification. The feature vector, $z = (z_1, \dots .., z_K)$ 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, $d \in \{1, \dots, N_d\}$ 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.

In many processes the input in the source will change describing a discrete state sequence; In this case the observation sequence can have time-varying characteristics. Different states can have different probability density distributions for the output signal. A better optimized 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 aids application speech analysis process.

An ANFIS system could be used with fuzzy neural network model. Such framework makes the ANFIS modeling more efficient and less dependent on expert knowledge and therefore facilitates learning and adjustment. In the first layer, all the nodes are adaptive nodes. The outputs of layer 1 are the fuzzy membership grade of the inputs, which are given by [26]:

$$O_i^1 = \mu_{Ai}(x), \; O_i^1 = \mu_{B(i-1)}(x), \tag{16}$$

Where x is the input to node i and A_i and B_{i-1} is the linguistic label associated with this function.

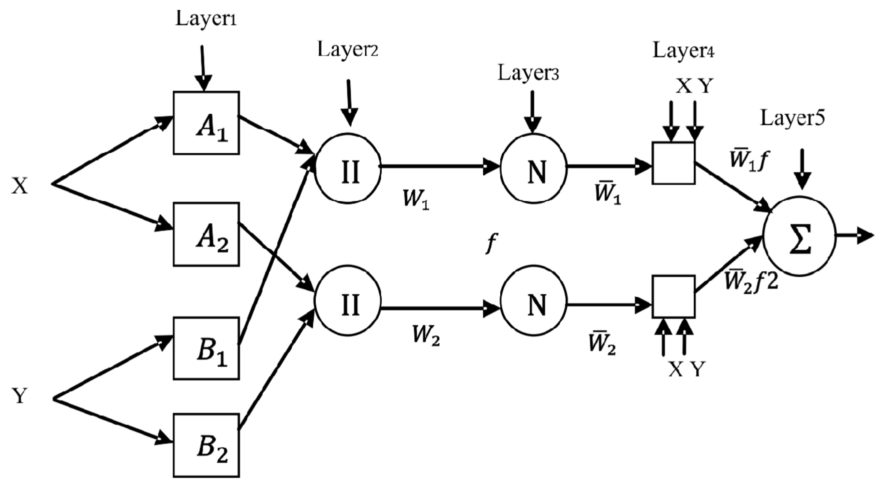

Figure 2: ANFIS Architecture

O_i^1 is the i th output of layer 1, $\mu_{Ai}(x)$ and $\mu_{B(i-2)}(x)$ and are type A and type B arbitrary fuzzy membership functions of nodes i and $i-2$ respectively. In the second and third layer, the nodes are fixed nodes. They are labeled M and N respectively, indicating they perform as a simple multiplier. The outputs of these layers can be represented as

$$O_1^2 = w_i = \mu_{Ai}(x)\mu_{Bi}(x) \tag{17}$$

$$O_1^3 = \overline{w_i} = \frac{w_i}{w_i + w_{i+1}} \tag{18}$$

Where $\overline{w_i}$ is the normalized firing strengths. In the fourth layer, the nodes are adaptive nodes. The output of each node in this layer is simply 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 by:

$$O_1^4 = \overline{w_i} f_i = \overline{w_i}(p_i x + q_{iy} + r_i) \tag{19}$$

in which f_i is the firing rate, p_i is the x scale, q_i is the y scale, and r_i is the bias for i th node. In the fifth layer, there is only one single fixed node that performs the summation of all incoming speech signals:

$$O_1^5 = \sum_{i=1}^{2} \overline{w_i} f_i = \sum_{i=1}^{2} \frac{w_i f_i}{w_i + w_{i+1}} \tag{20}$$

From the proposed ANFIS architecture in Figure 2. it is observed that given the values of basis parameters, the overall output can be expressed as a linear combinations of the consequent output σ, The task of the learning algorithm for this architecture is to tune all the above mentioned modifiable parameters to make the ANFIS output match the training data. When the premise parameters of the membership function are fixed, the output of the ANFIS model can be written as:

$$\sigma = \frac{w_1}{w1+w_2} f_1 + \frac{w_2}{w1+w_2} f_2 => \overline{w_1} f_1 + \overline{w_2} f_2$$

$$=> (\overline{w_1}x)p_1 + (\overline{w_1}y)q_1 + (\overline{w_1})r_1 + (\overline{w_2}x)p_2 + (\overline{w_2}y)q_2 + (\overline{w_2})r_2 \tag{21}$$

Which is linear in the consequent parameters $(p_1, q_1, r_1, p_2, q_2 and\ r_2)$. When the premise parameters are fixed the least squares method is used easily to identify the optimal values of these parameters after adjustment of ANFIS weights. When the premise parameters are not fixed, the search space becomes larger and the convergence of the training becomes slower.

Proposed optimized FA-ANFIS Network speech signal classification

After finding fitness value of the selected features of the speech signal using Firefly algorithm, an ANFIS classifier that takes an assessment based on the best fitness values of selected features of their distance measure, the speech signal is classified. In speech signal classification, the feature vector as a given input to the classifier. Since the pure speech signal is computationally demanding to use directly, the relevant information is selected from the speech signal.

4. EXPERIMENTATION RESULTS

The experimentation process used the RWCP- sound scene database (SSD) [27] for performance evaluation of the proposed speech enhancement process in hearing aids application. The database used in the proposed method contains sounds in real acoustical environments. The proposed methods are simulated on MATLAB, the simulation results are used for evaluation of quality of enhanced speech signal. Speech intelligibility and speech quality measures reflect the true performance of any speech enhancement algorithm as well as Noise reduction Technique. Speech Quality assessment of the proposed method is evaluated by using different quality measures. This quality measure results of the proposed techniques are compared with existing and. The quality measures such as Segmental Signal to Noise ratio (SNR_{seg}) which is required for better quality assessment for 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. The above mentioned techniques of speech analysis were applied to the noisy speech input and the performance parameters were evaluated as below.

SNR_{seg}

Segmental Signal-to-Noise Ratio (SNR_{seg}), instead of working on the whole signal, calculates the average of the SNR values of short segments (t is given by:

$$SNR_{Seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \sum_{i=Nm}^{Nm+N-1} \left(\frac{\sum_{1=1}^{N} x^2(i)}{\sum_{1=1}^{N} \left(x(i) - y(i)\right)^2}\right) \tag{22}$$

Where N and M are the segment length and the number of segments respectively.

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 A:

$$PESQ = a_0 - a_1 . D - a_2 . A \tag{24}$$

Where, $a_0 = 0.1$, $a_1 = 0.1$, and $a_2 = 0.0309$

The clean and noise signals were processed by the Fast ICA noise reduction algorithm, improved DWT feature extraction and optimized FA-ANFIS classifier to simulate the receiving frequency characteristics of hearing aids application. Then the noise signals were added to the clean speech at different SNR_{seg}, and PESQ levels respectively. This proposed work used to select three types of background noise: white noise, babble noise and car noise. The noisy speech signals were enhanced by different speech enhancement algorithm like Spectral subtraction, wiener filter, genetic SVD, Hybrid PCA-ANFIS Modified Bayesian-ANFIS and optimized FA-ANFIS which cover the major classes of noise reduction in the speech analysis process.

Figures 3, 4, 5, 6. shows that the results across different speech processing algorithms under three background noises at SNR and PESQ levels. The Figures shows that the overall performance of the proposed method, which is high compare to existing methods.

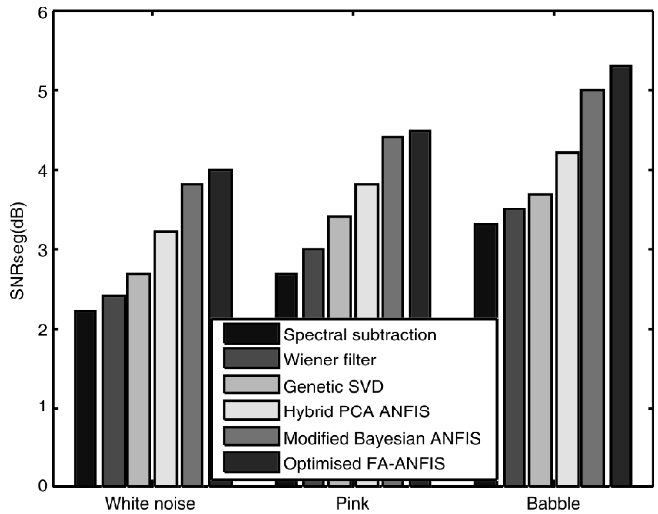

Figure 3: SNR_seg 3. vs Types of Noises

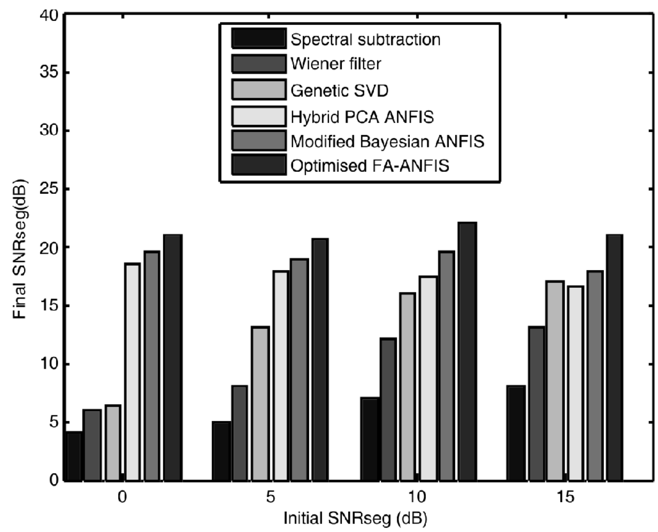

Figure 4: The Comparison results of using different algorithms

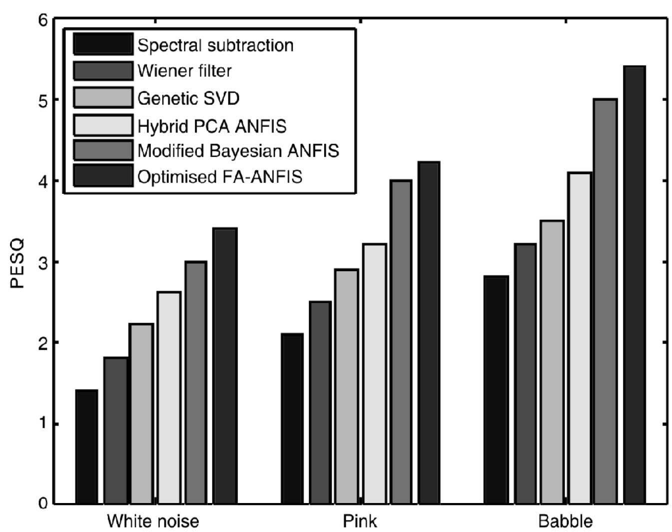

Figure 5: PESQ vs. Types Noises

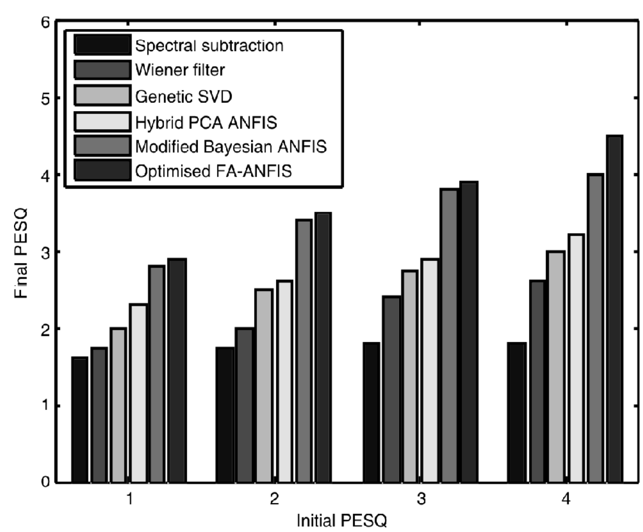

Figure 6: The Comparison results of using different algorithms

5. CONCLUSION

In this work proposed optimized classification method for classifies the background noises and enhanced speech signals. There are different no. of background noises present in speech signals, which are affect the speech intelligibility and quality of the speech signals in hearing aids application. To increase the performance of the hearing aids application, the enhanced speech signals are correctly classified with the use of proposed optimized FA- 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 optimized classification algorithm. In this work used proposed method speech enhancement method such as Fast Independent Component Analysis (Fast ICA) for reducing the background noise and the Improved Discrete Wavelet Transforms could effectively extract the features from the speech signal and AANOVA method used for select the significant features for efficient classification results, finally the optimized FA-ANFIS method classify the enhanced speech signal for efficient manner to increase the speech intelligibility and quality in the hearing aids.

REFERENCES

[1] M. Brandstein and D. Ward, "Microphone Arrays: Signal Processing Techniques and Applications," Digital Signal Processing, 2001, Springer.

[2] J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, pp. 471–472, Oct. 1978.

[3] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp. 113–120, Apr. 1979.

[4] Virag, N. (1999). Single channel speech enhancement based on masking properties of the human auditory system. Speech and Audio Processing, IEEE Transactions on, 7(2), 126-137.

[5] Khemili, I., & Chouchane, M. (2005). Detection of rolling element bearing defects by adaptive filtering. European Journal of Mechanics-A/Solids, 24(2), 293-303.

[6] Hardwick, J., Yoo, C. D., & Lim, J. S. (1993, April). Speech enhancement using the dual excitation speech model. In Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on (Vol. 2, pp. 367-370). IEEE.

[7] J. S. Lim, A. V. Oppenheim, and L. D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, pp. 354–358, Aug. 1978.

[8] De Cheveigné, A. (1993). Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time domain cancellation model of auditory processing. The Journal of the Acoustical Society of America, 93(6), 3271-3290.

[9] Arons, B. (2008). A Review of The Cocktail Party Effect.

[10] S. F. Boll and D. C. Pulsipher, "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, pp. 752–753, Dec. 1980.

[11] Daqrouq, K., Abu-Isbeih, I. N., & Alfauri, M. (2009, March). Speech signal enhancement using neural network and wavelet transform. In Systems, Signals and Devices, 2009. SSD'09. 6th International Multi-Conference on (pp. 1-6). IEEE.

[12] Abd El-Fattah, M. A., Dessouky, M. I., Diab, S. M., & Abd El-Samie, F. E. S. (2008). Speech enhancement using an adaptive wiener filtering approach. progress in electromagnetics research M,4, 167-184.

[13] Mousa, A., Qados, M., & Bader, S. (2012). Speech signal enhancement using adaptive noise cancellation techniques. Canadian Journal on Electrical and Electronics Engineering, 3(7), 375-383.

[14] Ravi, B., & Kumar, T. K. (2013). Speech Enhancement using Kernel and Normalized Kernel Affine Projection Algorithm. arXiv preprint arXiv:1309.2359.

[15] Wolfe, P. J., & Godsill, S. J. (2002). Bayesian estimation of time-frequency coefficients for audio signal enhancement. In Advances in Neural Information Processing Systems (pp. 1197-1204).

[16] C. H. Taal and R. Heusdens, "A low-complexity spectrotemporal based perceptual model," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 153–156.

[17] R. C. Hendriks, R. Heusdens, and J. Jensen, “MMSE based noise PSD tracking with low complexity,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 4266–4269.

[18] Levitt H. New trends—Digital hearing aids: Past, present, and future. In: Practical Hearing Aid Selection and Fitting. Veterans Administration Rehabilitation Research and Development Service 1997; Monograph 001, p. xi-iii.

[19] Dillon H, Lovegrove R. Single microphone noise reduction systems for hearing aids: A review and an evaluation. In: Studebaker GA, Hochberg I (eds.). Acoustical factors affecting hearing aid performance, second edition. Boston: Allyn and Bacon, 1993.

[20] A. Hyvarinen, J. Karhunen, E. Oja: *Independent Component Analysis*, John Wiley & Sons, New York etc., 2001.

[21] Louis, A.K., Maass, D. and Rieder, A. (1997) Wavelets-Theory and Applications. Wiley, Hoboken.

[22] Avci, E. (2007) A New Optimum Feature Extraction and Classification Method for Speaker Recognition: GWPNN. *Expert System with Applications*, 32, 485-498.

[23] Burrus, C.S., Gopinath, R.A. and Guo, H. (1998) Introduction to Wavelet and Wavelet Transforms. Prentice Hall, New Jersey.

[24] Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. Data Classification: Algorithms and Applications, 37.

[25] X. S. Yang, Nature-Inspired Metaheuristic Algorithms, Luniver Press, UK, (2008).

[26] Jang, Jyh-Shing Roger. “ANFIS: adaptive-network-based fuzzy inference system.” Systems, Man and Cybernetics, IEEE Transactions on 23, no. 3 (1993): 665-685.

[27] Dataset from :http://research.nii.ac.jp/src/en/RWCP-SSD.html.