National Conference on
CASH LESS TRANSACTIONS USING
COMPUTER AND COMMUNICATIONS
www.worldconferences.org
www.ijsetr.com

Speech Signal Enhancement using GS Optimization

ANOOP. V[1], P.V. RAO[2]
[1]Dept of ECE, VTU, Belgaum, India, E-mail: vanoop@gmail.com.
[2]Professor, Dept of ECE, Vignana Bharathi Institute of Technology, Hyderabad, TS, India.

Abstract: Transmission quality of the signal may possibly considerably deteriorate from the existence through the noise disturbance. In case of interference, the transmitted signal denoising is usually highly attractive for that reason in the speech signal enhancement process. In our proposed signal enhancement process, the desired signal quality will be enhancing by means of employing a series of processes particularly, AMS Feature Extraction, Signal Enhancement, GSO optimization and finally analyzing its performance. Initially, the input signal distracted with noise will be taken and for which the AMS features were extracted. At the signal enhancement stage, a new signal will be obtained as a product of the feature extracted and the original signal. Then the signal obtained after enhancement process is again optimized by means of passing through the GSO optimization algorithm. Finally the SNR values will be evaluated for all the signal outputs to analyze its performance. The proposed technique is evaluated in the working platform of MATLAB and the results were analyzed.

Keywords: Speech Enhancement GSO, AMS Feature Extraction, Optimization SNR.

I. INTRODUCTION

Speech enhancement refers to the improvement in the quality or intelligibility of a speech signal and the reversal of degradations that have corrupted it. Quality is a subjective measure which reflects on the pleasantness of the speech or on the amount of effort needed to understand the speech material. Intelligibility is an objective measure which signifies the amount of speech material correctly understood. The main objective of Speech Enhancement is to enhance the speech signal to obtain a clean signal with higher quality. Such system has been widely used in long distance telephony applications [1]. Speech enhancement techniques can be classified into, single channel, dual channel or multi-channel enhancement. Although the performance of multichannel speech enhancement is better than that of single channel enhancement, the single channel speech enhancement is still a significant field of research interest because of its simple implementation and ease of computation [3]. Digital signal processing (DSP) techniques for speech enhancement include spectral subtraction, adaptive filtering and suppression of non harmonic frequencies. Most of these techniques either require a second microphone to provide the noise reference or require that the characteristics of noise be relatively stationary [5]. Spectral subtraction has been widely used for enhancing speech because of its simplicity and ease of implementation in single channel systems but it suffers from the production of musical noise after enhancement and is one of its major drawbacks [2]. Although the MMSE-STSA method gives an estimated speech signal with less musical noise, it requires more complicated computations, for example, the solution required to calculate the modified Bessel function. Moreover, as pointed out by some researchers, real speech histograms do not fit to Rayleigh function employed [7]. There is a strong need to improve the quality of the speech signal in noisy conditions by developing speech enhancement algorithms to minimize the effect of background noise. In the gain adjustment process the quality of the speech signal in noisy environment is improved by automatically adjusting the output level when the background noise exceeds the noise masking threshold [6]. Further, there are two important points often required to be considered in speech enhancement applications; eliminating the undesired noise from the speech to improve the Signal-to-Noise Ratio (SNR), and retrieving the quality of the original speech signal which leads to improvement of the speech intelligibility [4].

II. PROPOSED METHODOLOGY

The proposed method is mainly to enhance the speech signals. The innovative technique involves the process of optimization of the signal input by passing through the Group Search Optimization (GSO) algorithm after taking the AMS (Amplitude Magnitude Spectrum) features. Initially the desired signal added with noisy signals will be taken and are extracted with the Amplitude Magnitude spectrum features.

A. Platform Associated with proposed methodology

The proposed framework involves the following steps,

1. Speech signal synthesis
2. AMS Feature Extraction (Level Adjustment, Frame Conversion, FFT, Hamming Window, Zero Padding)
3. Signal Enhancement
4. GSO Optimization
5. Signal performance Analysis

1. Speech Signal Synthesis

The signal input will be taken and for which the AMS features were extracted. Let $q(m)$ be the original signal defined for all m . The original desired signal usually contains some noise added with it. It is important to denoise the signal in order to obtain proper the exact transmitted signal. Let the noise added to the original signal is given as follows,

$$input\ \ signal = desired\ \ signal + noise \tag{1}$$

$$q(m) = r_m + \sigma n(m) \tag{2}$$

Where, $n(m)$ - noisy signal.

For white noise, the noisy signal $n(m)$ will be normally distributed between (0,1).

2. AMS Feature Extraction

In the AMS Feature extraction stage, the signal input is first taken and for which the level is adjusted along with this the process of sampling takes place. Where, the signal is resized to some desired size. Then the signal is to be divided into number of frames, i.e., signal fragmentation. Then the fast Fourier transformation will be taken for the frame segregated signal.

FFT (Fast Fourier Transform): The fast Fourier transform (FFT) is a discrete Fourier transform algorithm which reduces the quantity of computations required for M points. Let us considering the number of computations required be of $2M^2$, and with the FFT transform, we can reduce the number of computation requirement from $2M^2$ to $2M\log(M)$. The Fast Fourier transform of $q(m)$ is to be evaluated at time m . Fast Fourier transformation algorithms can generally fall into two classes:

- Decimation in Time
- Decimation in Frequency

The basic idea is to break up a transform of length M into two transforms of length $\frac{M}{2}$ using the identity given as below,

$$\sum_{m=1}^{M-1} q_m e^{-\frac{2\Pi jmp}{M}} = \sum_{m=0}^{\frac{M}{2}-1} q_{2m} e^{-\frac{2\Pi j(2m)p}{M}} + \sum_{m=0}^{\frac{M}{2}-1} q_{2m} e^{-\frac{2\Pi j(2m+1)p}{M}} \tag{3}$$

$$= \sum_{m=0}^{\frac{M}{2}-1} q_m^{even} e^{-\frac{2\Pi jmp}{M/2}} + e^{-\frac{2\Pi jmp}{M}} \sum_{m=0}^{\frac{M}{2}-1} q_m^{even} e^{-\frac{2\Pi jmp}{M/2}} \tag{4}$$

Based on the FFT outputs, some threshold will be set so that the sub band of signals will be obtained by splitting the signals. Then the time domain is calculated for each sub bands.

Hamming Windowing: To improve the filter properties, a smoothing window (i.e., Hamming window) is applied when the frequency is changed continuously. Thus it is important to design a window reliable with the desired time and frequency. In this paper the filtering process is made with Hamming Windowing.

The Hamming Window function is given as for duration m for M number of samples as,

$$W_{Hamm}(m) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\Pi m}{M-1}\right), & 0 \le m \le M-1 \\ 0, & otherwise \end{cases} \tag{5}$$

Where, M - Order of filter.

The above equation can also be written as follows:

$$W_{Hamm}(m) = \begin{cases} 0, m < 0, m > M \\ \neq 0,\ 0 \le m \le M-1 \end{cases} \tag{6}$$

The aim of the pre-filtering is to improve the accuracy of the frequency determination. After smoothening of the signal, it is to be zero padded i.e., adding zeros to the signal outputs so that the signal at time domain could increase its length.

Zero Padding: The windowed signal is then zero padded, where the signals at time domain will be adjusted to the length equal to the other signals. Zero padding before the frequency transformation results in smoother spectra, we ensure that adjoining frequency values of the inputs will be close to each other. Finally, the AMS features will be extracted by the above series of processes and are given to signal enhancement stage.

3. Signal Enhancement

In signal enhancement stage, the feature extracted output will be multiplied with the original signal and the product is utilized to find the distance. Let the signal obtained after extracting AMS features be W_i , which is to be multiplied with the noised signal to obtain a new solution Z_i

$$Z_i = q(m) \times W_i \tag{7}$$

Where, $q(m)$ - Input signal W_i - Feature Extracted output.

To denoise the signal and to enhance the desired signal, the AMS feature extracted signal is to be GSO optimized. In GSO optimization, the Feature Extracted output (W_i) will be made as the initial solution. On further processing, the original signal will be replaced by the product obtained from above equation in case of minimum fitness value. The minimum fitness value here is the minimum Euclidean distance between the two solutions.

4. Group Search Optimization (GSO)

The technique of Group Search Optimization (GSO) is employed here to enhance the speech signal after the extraction of AMS features. The initial set of solution to the algorithm is termed as group and the members of the group were categorized as, Producer, Scrounger and Ranger. Producers are the members of group which will go in search of finest resources. The members of scrounger find the association between the resources, which was discovered by the producers. Rangers are the members that generate activities in an irrelevant manner along with searching in a prepared means, so that efficient finding of resources may be accomplished.

GSO algorithm: Let us considering the initial solution is the final AMS feature extracted signal output. The initial solution to the optimization process is given in the form as,

$$W_i = \begin{bmatrix} w_{11} & w_{12}......w_{1m} \\ w_{21} & w_{22}......w_{2m} \\ w_{n1} & w_{n2}......w_{nm} \end{bmatrix} \tag{8}$$

Search Direction vector: The search direction of the k^{th} member at searching space h in the k-dimensional search space is given as,

$$V_k^h\left(\xi_k^h\right) = (v_{k1}^h, v_{k2}^h, ..., v_{kl}^h) \in R^l \tag{9}$$

The search direction vector can be calculated by means of the below equations (10), (11) and (12) by the means of polar and Cartesian Coordinate Transformations as,

$$v_{k1}^h = \prod_{p=1}^{l-1} \cos(\xi_{kp}^r) \tag{10}$$

$$v_{kr}^h = \sin(\xi_{k(r-1)}^h) \prod_{p=r}^{l-1} \cos(\xi_{kp}^h) \ \ for \ (r = 2,3,...,l-1) \tag{11}$$

$$v_{kl}^h = \sin(\xi_{k(l-1)}^h) \tag{12}$$

Head angle:

The head angle is utilized in the generation of best resources, as the producer will rotates its head inorder to search the finest resource. The angle to which the producer should rotate is denoted as the head angle. The Head angle representation is given as,

$$\xi_k^h = (\xi_{k1}^h, \xi_{k2}^h, ..., \xi_{k(l-1)}^h) \tag{13}$$

Fitness Function: The fitness function is given as the minimum of Euclidean distance and is given as,

$$F_f = \min(E_d) \tag{14}$$

$$E_d = \sqrt{\sum_{i=m=1}^{M} (Z_i - r_m)^2} \tag{15}$$

Where, Z_i - New Signal generated r_m - Desired signal

B. Representation for head angle rotation at varying angles of producer

Also, the position of rotating the head angle is mainly made taking into account with three categories of rotation particularly at zero degrees, at right hand side hypercube, at left hand side hypercube

The representation for head angle rotation at varying angles of producer is given as below,

At zero degrees:

$$W_Z = W_{pr}^h + s_1 t_{\max} V_{pr}^h(\xi^h) \tag{16}$$

At right hand side hypercube:

$$W_R = W_{pr}^h + s_1 t_{\max} V_{pr}^h\left(\xi^h + s_2 \frac{\delta_{\max}}{2}\right) \tag{17}$$

At left hand side hypercube:

$$W_L = W_{pr}^h + s_1 t_{\max} V_{pr}^h\left(\xi^h - s_2 \frac{\delta_{\max}}{2}\right) \tag{18}$$

In the above equations, $s_1 \in R^1$ and $s_2 \in R^{l-1}$

Where,

s_1 - Normally distributed random number (i.e., $Mean = 0$ and $SD = 1$(Standard Deviation))

s_1 - Uniformly sequenced random number between range [0, 1]

$\delta_{\max}$ - maximum degree of rotating angle at the hypercube

The best resource will be generated by the producer by randomly generating the rotating angle, which is given below as,

$$W^{h+1} = \xi^h + s_2 \gamma_{\max} \tag{19}$$

In the above equation, $\gamma_{\max} \in R^1$, where $\gamma_{\max}$ is the maximum turning angle.

For minimum turning angle (occurs in case of not obtaining best resource), substituting $\gamma_{\max} = 0$ in the above equation,

$$W^{h+x} = \xi^h \tag{20}$$

Where, $x \in R^1$ is a constant

The association between the resources was found out by the scrounger which is given as below,

$$W^{h+1} = W_k^h + s_3 \circ (W_{pr}^h - W_k^h) \tag{21}$$

Where, $s_1 \in R^l$ is the uniform random sequence between range [0, 1] and "$\circ$"- operator referring Hadamard product (or) Schur product.

Finally, the rest of the group members (rangers) will be dispersed from their current positions based on distance measure generated through the randomly generated head angle. The distance measure is given as below,

$$t_s = x . s_1 . t_{\max} \tag{22}$$

The rangers finds the resources by means of the below equation,

$$W_k^{h+1} = W_k^{\ h} + t_k\, V_k^h\left(\xi^{h+1}\right) \tag{23}$$

At last, the process of evaluation takes place, in which the fitness function is again obtained and the solution with better fitness function will be replaced by the current solution. The process of finding best resource continues until a best one is obtained as fittest resource.

In the final stage, the original signal is enhanced. The enhanced signal is obtained through the GSO optimization, where the input solution with minimum Euclidean distance is replaced by the desired signal. So that the output obtained solution will be enhanced.

Signal performance Analysis:

The enhanced signal will be analyzed by calculating the signal to noise ratio values of both the final enhanced signal and the desired signal. SNR can be defined as the ratio of signal power to the noise power. The SNR value can be obtained through the below equation as,

$$SNR = \frac{Signal\ \ Power}{Noise\ \ Power} \tag{24}$$

The above equation can also be written as follows, considering the input signal to be $x(m)$ and the noise signal is of n :

$$SNR = \frac{P_{x(m)}}{P_n} \tag{25}$$

Where, $P_{x(m)}$ - Signal Power P_n -Noise Power

III. RESULTS AND DISCUSSION

The outcomes attained and the resulting debates based on them are detailed in this section. The investigational set up and replication outcomes are thrashed out below. The database is used for obtaining noise sounds including suburban train noise, babble, car, exhibition hall, restaurant, street, and airport and train-station noise. The sentences were initially sampled at 25 kHz and then down sampled to 8 kHz and noise segment is synthetically supplemented to the speech signal. The proposed technique for speech enhancement is implemented in a system having 8 GB RAM with 32 bit operating system having i5 Processor using MATLAB Version 2014. The signal power is plotted for a frequency range between 0 to 2.5 KHz. The input signal, noisy signal and the de-noised signal are shown in figures.

Table 1. SNR Values of different noises at Different Levels.

Noise Types	Noise Levels		
	5dB	10dB	15dB
Babble	18.784	24.567	29.561
Car	19.002	23.347	29.689
Exhibition	18.564	22.989	29.456
Street	18.876	23.765	28.888
Restaurant	19.234	24.056	29.236

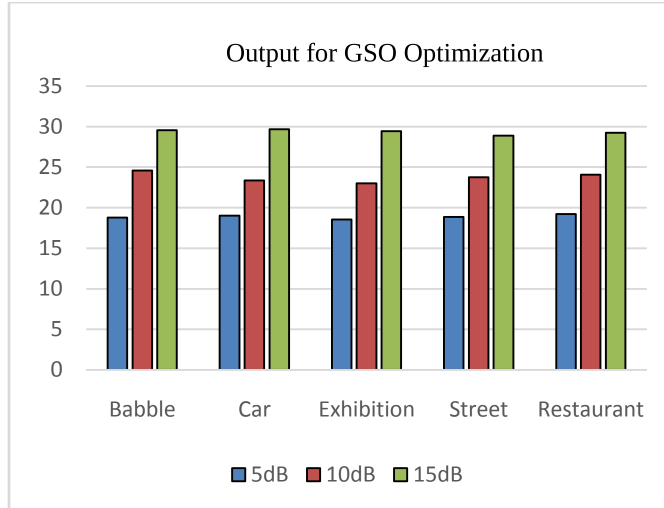

Figure1.SNR values for Proposed Babble, Car, Exhibition, Street, and Restaurant noises of 5db, 10dB and 15dB using GSO Optimization.

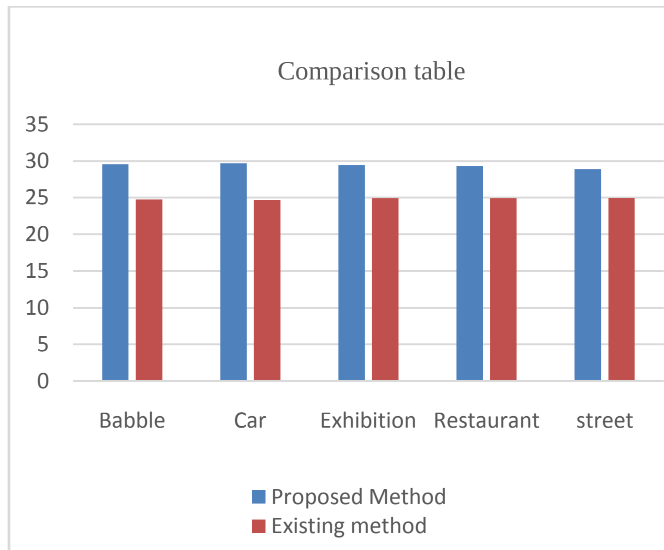

Figure2.Comparison of SNR value15 dB for Proposed and existing method.

IV. CONCLUSION

Denoising associated with transmitted signal consists of the transmission to be clearly represented and filtered, to remove out all noises and artifacts from the signal. While mainly because all the signal outputs attained are influenced by disturbance. The actual disturbance degrades the exactness in addition to accuracy of evaluation, and decreases the detection restriction on the strategy. In this work, the signal will be obtained and are extracted with AMS features and enhanced to improve its quality. The signal enhancement will be made more appropriate with the optimization process of GSO algorithm. Then finally, SNR values were evaluated for all the signal outputs and are compared with the original signal to evaluate the performance of the signals. The resultant outputs for both the proposed and existing systems were obtained and are analyzed to show the efficiency of our proposed technique.

V. REFERENCE

[1] Pankaj Bactor1, Anil Garg, “Different Techniques for the Enhancement of the Intelligibility of a Speech Signal,” International Journal of Engineering Research and Development, Vol. 2, Pp. 57-64, 2012.
[2] Sudeep Surendran and T. Kishore Kumar, "Perceptual Subspace Speech Enhancement with Variance Normalization", Procedia Computer Science, No. 54, pp. 818 – 828, 2015.
[3] Pinki, Sahil Gupta, " Speech Enhancement using Spectral Subtraction-type Algorithms: A survey on Comparison", International Journal of Engineering and Computer Science, Vol. 4, Pp. 14873-14878, 2015.
[4] Amin Zehtabian, Hamid Hassanpour, Shahrokh Zehtabian, Vicente Zarzoso, "A Novel Speech Enhancement Approach Based on Singular Value Decomposition and Genetic Algorithm", In Proceedings of 2nd International Conference on Soft Computing and Pattern Recognition, Pp. 430-435, 2010.
[5] Hossein Sameti, Hamid Sheikhzadeh, Li Deng, and Robert L. Brennan, "HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise", IEEE Transactions on Speech And Audio Processing, Vol. 6, No. 5, 1998.
[6] Premananda B S and Dr. Uma B V, "Speech Enhancement Algorithm to Reduce the Effect of Background Noise in Mobile Phones", International Journal of Wireless & Mobile Networks (IJWMN), Vol. 5, No. 1, 2013.
[7] Arata Kawamura, Weerawut Thanhikam, and Youji Iiguni, "Single Channel Speech Enhancement Techniques in Spectral Domain", 2012.
[8] Anil Garg, O.P. Sahu, “Cuckoo search based optimal mask generation for noise suppression and enhancement of speech signal,” Journal of King Saud University - Computer and Information Sciences, Vol. 27, Pp. 269-277, 2015.
[9] Samira Mavaddaty, Seyed Mohammad Ahadi, Sanaz Seyedin, "Novel Speech Enhancement Method by Learnable Sparse and Low-Rank Decomposition and Domain Adaptation", Speech Communication, Vol. 76, Pp. 42-60, 2016.
[10] Mohamed anouar Ben messaoud, Aïcha Bouzid, Noureddine Ellouze, "Speech enhancement based on wavelet packet of an improved principal component analysis", Computer Speech & Language, Vol. 35, Pp. 58-72, 2016.
[11] Ji-Won Cho, Hyung-Min Park, "Independent vector analysis followed by HMM-based feature enhancement for robust speech recognition", Signal Processing, Vol. 120, Pp. 200-208, 2016.
[12] Philip Harding, Ben Milner, "Reconstruction-based speech enhancement from robust acoustic features", Speech Communication, Vol. 75, Pages 62-75, 2015.