International Journal of Engineering Innovation & Research
Volume 2, Issue 6, ISSN: 2277 – 5668

Speech Signal Quality Improvement Using Cuckoo Search Algorithm

Anoop V.
Email: vanoop@gmail.com

Dr. P. V. Rao
Email: pachararao@rediffmail.com

***Abstract* – Speech signal quality improvement attempt to attenuate ambient noise to reduce listener fatigue and improve intelligibility. This requires maintaining speech cues and ensuring natural sounding residual noise that does not distract the listener. In this paper, an effective noise suppression technique for enhancement of speech signals using optimized window is proposed. Initially, the noisy speech signal is broken down into various time-frequency (TF) units and the features are extracted by finding out the Amplitude Magnitude Spectrogram (AMS). The signals are then classified based on quality ratio into different classes to generate the initial set of solutions. Subsequently, ideal window for each class is generated based on Cuckoo search algorithm. Subsequently, in the waveform synthesis stage, filtered waveforms are windowed and then multiplied by the ideal window value and summed up to get the enhanced target signal. The results obtained proved the effectiveness of the proposed technique and is able to suppress noise and improve the quality of the speech signal.**

***Keywords* – Speech Signal, Speech Enhancement, Cuckoo Search, AMS, SNR, and Optimal Mask.**

I. Introduction

Speech enhancement (SE) aims to improve the performance of speech communication systems in noisy environments [1]. SE applied in, for example, to a mobile radio communication system, a speech to text system, a speech recognition system, a set of low quality recordings, or to improve the performance of aids for the hearing impaired [5, 9]. SE is a classical problem in signal processing. Hearing aid users often have great difficulty understanding speech in a noisy background. They typically require a signal-to-noise ratio SNR of about 5–10 dB higher than normal hearing listeners to achieve the same level of speech understanding [11]. Therefore, several single and multi-microphone noise reduction strategies have been developed for modern hearing aids. Multi-microphone noise reduction systems are able to exploit spatial in addition to spectral information and are hence preferred to single-microphone systems [12]. The enhancement of a desired speech signal in the presence of stationary noise using an array of microphones has been studied for several years [13]. Algorithms for the enhancement of acoustically distorted speech signals are used for a variety of applications such as hands-free devices, mobile phones or hearing aids. Commonly used systems for (single channel) SE achieve a suppression of disturbing background noise, e.g., [14] but do not (notably) reduce speech distortions due to room reverberation [15]. In many speech enhancement and noise reduction algorithms, the decision is based on the a priori SNR [2], and the classic algorithms like spectral subtraction, Wiener filtering, and maximum likelihood, can be formulated as a function of this a priori SNR [3]. In real-time applications, the a priori SNR estimation is useful, but in the ideal situation the local SNR is preferable instead of the a priori SNR [6]. For example, Ephraim and Malah used the decision directed approach for signal-to noise ratio estimation by using the weighted average of the past SNR estimate and the present SNR estimate [8].

The types of distortion introduced by SE algorithms can be broadly divided into two categories: the distortions that affect the speech signal itself and the distortions that affect the background noise (called noise distortion). Of these two types of distortion, listeners seem to be influenced the most by the speech distortion when making judgments of overall quality [16]. Unfortunately no objective measure currently exists that correlates high with either type of distortion or with the overall quality of speech enhanced by noise suppression algorithms [17]. To enhance the quality and intelligibility of noisy speech, research in SE [18] has focused on better modeling of the speech and noise PDFs, the way the noise contaminates the clean speech, the type of noise source, etc. The most common distortion in speech is due to additive noise, which is independent of the clean speech [19]. The SE algorithms explored can be grouped into two major classes [20]: 1) the class based on hidden Markov model (HMM) [21], and 2) the class based on transformation of signals, such as MMSE estimation [22], spectral subtraction [18] and subspace based methods [20], [23]. Ephraim and Van Trees introduced subspace based approach [23]. Previously, different noise reduction methods have been proposed [10]. Wavelet-based techniques using coefficient thresholding approaches have also been applied for speech enhancement [4]. As alternative to these traditional techniques is studying speech as a nonlinear dynamical system [7].

Although subjective evaluation of SE algorithms is often accurate and reliable provided it is performed under stringiest conditions (e.g., sizeable listener panel, inclusion of anchor conditions, etc. [24]), it is costly and time consuming. For that reason, much effort has been placed on developing objective measures that would predict speech quality with high correlation. Many objective speech quality measures have been proposed in the past to predict the subjective quality of speech [24]. Most of these measures, however, were developed for the purpose of evaluating the distortions introduced by speech codecs and/or communication channels [25]. The quantization and other types of distortions introduced by waveform and linear predictive coding (LPC)-based speech coders [e.g., code excited linear prediction (CELP)]; however, are different from those introduced by SE algorithms [17]. The a posteriori and a priori SNRs are main function for computing gain function using modified decision-directed approach [10]. The gain function used in ideal binary window for computational auditory scene analysis is identical to the gain function of the Maximum *a posterior* (MAP) estimators [26]. Although, speech produced in the presence of noise called "Lombard speech" has been found to be easily understandable than speech produced during silence [27]. In previous studies, large gain in intelligibility can be obtained by multiplying the noisy signal with the ideal binary window signal, even at extremely low (-5, -10 dB) SNR levels [28, 29]. The generation of binary window with the help of Bayesian classifier technique that is lazy classification technique. Since the classification with the lazy classifier, the generation of binary window will not be ideal one. If the binary window is not ideal one, it will affect the performance of the speech enhancement. This paper presents ideal window generation using cuckoo search algorithm [30] for speech enhancement to improve the SNR and thus intelligibility. The proposed algorithm optimizes the windowing parameters in order to suppress the noise effectively for enhancement of speech signal. A comparison and simulation results of our proposed method is better in terms of SNR than the Bayesian classifier technique.

The rest of the paper is organized as follows: a brief review of some of the literature works in speech enhancement domain is presented in Section II, Brief description of Cuckoo search algorithm and the cuckoo search based ideal window generation is explained in Section III. The experimental results and comparative analysis discussion is provided in Section IV. Finally, the conclusions are summed up in Section V.

II. Related Work

Literature presents lot of works for speech enhancement. Here, some of work related to the SE is reviewed in this section. Van den Bogaert T *et al.* [11] have presented speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids. They evaluated speech enhancement in binaural multi-microphone hearing aids by noise reduction algorithms based on the multichannel Wiener filter MWF and the MWF with partial noise estimate MWF-N. Both algorithms were specifically developed to combine noise reduction with the preservation of binaural cues. Objective and perceptual evaluations were performed with different speech-in-multitalker-babble configurations in two different acoustic environments. Adding the partial noise estimate to the MWF, done to improve the spatial awareness of the hearing aid user, reduces the amount of speech enhancement in a limited way. In some conditions the MWF-N even outperformed the MWF possibly due to an improved spatial release from windowing.

Lollmann, H.W. *et al.* [15] have presented a blind speech enhancement algorithm for the suppression of late reverberation and noise. They proposed a speech enhancement algorithm for the suppression of background noise and late reverberation without a priori knowledge. A generalized spectral weighting rule based on estimations for the spectral variances of the late reverberant speech and background noise was used for speech enhancement. This allowed suppressing speech distortions due to late room reflections without knowledge about the complete room impulse response. In contrast to existing methods, all needed quantities were estimated entire blindly from the reverberant and noisy speech signal. The algorithm had also a low signal delay which was important, e.g., for speech enhancement in mobile communication devices or hearing aids.

Achintya Kundu *et al.* [19] have presented GMM based bayesian approach to speech enhancement in signal transform domain. They considered a general linear model of signal degradation, by modeling the probability density function (PDF) of the clean signal using a Gaussian mixture model (GMM) and additive noise by a Gaussian PDF, they derived the minimum mean square error (MMSE) estimator. The derived MMSE estimator was non-linear and the linear MMSE estimator was shown to be a special case. For speech signal corrupted by independent additive noise, by modeling the joint PDF of time-domain speech samples of a speech frame using a GMM, they proposed a speech enhancement method based on the derived MMSE estimator. They also showed that the same estimator can be used for transform-domain speech enhancement.

Brady Laska *et al.* [31] have demonstrated that the intra-speech residual noise in RBPF enhanced speech results from under attenuation of noise in spectral troughs. Low-order full band TVAR models were insufficient to capture the large dynamic range of the wideband speech spectrum, while high-order models greatly increase complexity and were difficult to reliably estimate. In order to improve the noise reduction performance, they used the RBPF algorithm to enhance the speech DCT coefficients rather than the time for Wiener filtering due to the ability of the DCT to efficiently de-correlate speech. Operating in the transform domain allowed the statistical model approaches to offer low complexity and high levels of noise reduction, even for colored noises. Consequently, fluctuations of the noise signal spectrum from its expected value led to successive over and under-estimation of the true noise spectrum in a given frame, resulting in the well-known time-varying narrowband musical noise artifacts

Brady N. M. Laska *et al.* [32] have presented a particle filter approach to spectral amplitude speech enhancement. Spectral amplitudes were known to exhibit inter-frame dependencies and non-Gaussian statistics; however, incorporating these properties makes closed-form solutions intractable. Using the particle filter, the framework allowed the presented algorithm to model the speech spectral amplitudes as an autoregressive process with Laplace distributed excitation. All of the particle sampling distributions were constrained to take the measurement into account, improving sampling efficiency. In experiments using wideband speech and real recorded noise, the proposed algorithm variants were shown to offer natural-sounding output speech, with objective evaluation results that compare favorably to existing particle filter speech enhancement algorithms. The multiple model variant was found to improve inter-speech noise reduction, while the phase variant improves performance when the signal-to-noise ratio was low.

Gibak Kim and *et al.* [33] have proposed a traditional noise-suppression algorithm to improve speech quality, but not speech intelligibility. Motivated by prior intelligibility studies of speech synthesized using the ideal binary window, an algorithm is introduced that decomposes the input signal into time-frequency T-F units and makes binary decisions, based on a Bayesian classifier as to whether each T-F unit is dominated by the target or the window. Speech corrupted at low signal-to-noise ratio SNR levels −5 and 0 dB using different types of window is synthesized by this algorithm and presented to normal-hearing listeners for identification

Gibak Kim and Philipos C.Loizou [34] have focused on the development of an algorithm that can be optimized for a specific acoustic environment and improve speech intelligibility. The proposed method decomposes the input signal into time-frequency (T-F) units and makes binary decisions, based on a Bayesian classifier, as to whether each T-F unit is dominated by the target signal or the noise window. Target-dominated T-F units are retained while window-dominated T-F units are discarded. The Bayesian classifier is trained for each acoustic environment using an incremental approach that continuously updates the model parameters as more data become available. Listening experiments were conducted to assess the intelligibility of speech synthesized using the incrementally adapted models as a function of the number of training sentences.

Gibak Kim and *et al.* [35] have presented a theoretical framework that can be used to analyze potential factors that can influence the intelligibility of processed speech. More specifically, this framework focuses on the fine-grain analysis of the distortions introduced by speech enhancement algorithms. It is hypothesized that if these distortions are properly controlled, then large gains in intelligibility can be achieved. To test this hypothesis, intelligibility tests are conducted with human listeners in which we present processed speech with controlled speech distortions. The aim of these tests is to assess the perceptual effect of the various distortions that can be introduced by speech enhancement algorithms on speech intelligibility.

III. Speech Signal Quality Improvement Using Cuckoo Search Algorithm Based Ideal Window Generation

A. Cuckoo Search Algorithm

Cuckoo search (CS) is one of the latest optimization algorithms and was developed from the inspiration that the obligate brood parasitism of some cuckoo species laid their eggs in the nests of other host birds which is of other species. In Cuckoo Search, three idealized rules are considered which says that each cuckoo lays one egg at a time, and dump its egg in randomly chosen nest. The second rule states that best nests with high quality of eggs will carry over to the next generations and the third one says that the number of available host nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability in the range 0 to 1. In this case, the host bird can either throw the egg away or abandon the nest, and build a completely new nest. It is also assumed that a definite fraction of the nests are replaced by new nests. For a maximization problem, the quality or fitness of a solution can simply be proportional to the value of the objective function. The algorithm is based on the obligate brood parasitic behavior of some cuckoo species in combination with the Levy flight behavior of some birds and fruit flies.

B. Cuckoo Searches Based Ideal Window Generation

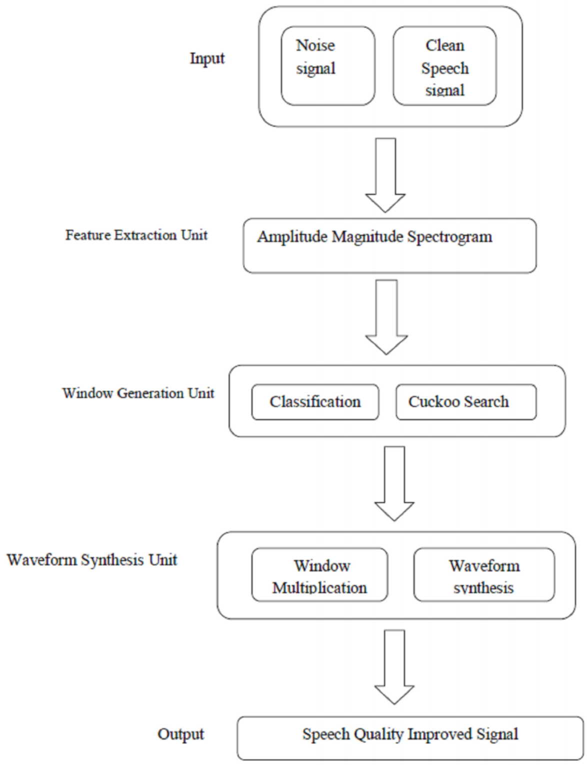

Fig.1. Block diagram of the proposed technique

The approach used in this paper for noise suppression and speech enhancement technique consists of three major units namely; Feature extraction unit, ideal window generation unit and waveform synthesis unit. Initially, the original and noise speech signal is given as input to extract features and subsequently, ideal window is generated with the use of cuckoo search. Subsequently, in the waveform synthesis unit, filtered waveforms are windowed and then multiplied by the ideal window value and summed up to get the improved quality signal. The block diagram of the proposed technique is shown in fig 1

C. Feature Extraction Unit

In this unit, features are extracted from the input speech corpus with the aid of the Amplitude Magnitude Spectrogram (AMS). The input speech signal will be a mixture of clean speech signal and the noisy signal. The input signal is initially processed by performing sampling, quantization and then, pre-emphasized to make the signal fit for further processing. Band-pass filter has the characteristics of passing the signals within the prescribed range of frequencies while attenuating other signals. Here, every channel is defined by the upper limit frequency U_i and the lower limit frequency L_i. After forming the channel bands, envelope of each band is calculated by the full wave rectification and subsequently, the envelope is decimated by a factor of 3 which is later segmented into overlapping segments of 128 samples of 32 ms with an overlap of 64 samples. The sampled signals obtained after the segmentation are Hanning windowed in order to remove unwanted signal components and get more sharper peaks. The windowed signals are initially zero-padded and taken Fourier transformed (256 point FFT) to obtain the modulation spectrum of each channel having frequency resolution of 15.6 Hz. Here, the signal is transformed into frequency domain signal decomposing the signal into its constituent frequencies. Let $x_0, x_1, \ldots, x_{N-1}$ represent the complex numbers and FFT is calculated as the formula given below

$$X_k = \sum_{k=0}^{N-1} x_n e^{-j2\pi k \frac{n}{N}}, \; where k = 0,1 \ldots N-1 \tag{1}$$

Hence, the modulation spectrum for all the 25 channels are obtained by the use of FFT and subsequently, every channel is then multiplied by fifteen triangular-shaped windows spaced uniformly across the 15.6–400 Hz range. Definition of triangular shaped windows is given below

$$w(n) = \frac{2}{N-1}\left(\frac{N-1}{2} - \left| n - \frac{N-1}{2} \right|\right) \tag{2}$$

Here N represents the width of the samples in the symmetrical window function and n is integer, with values varying from $0 < n < N-1$. All these are summed up to produce 15 modulation spectrum amplitudes and each of this represents the AMS feature vector. Let the feature vector is represented by $A_F(\lambda, \phi)$ where ϕ represents the time slot and λ represents the sub-band. Considering the small changes that may occur in the time and the frequency domains, we also take in the delta functions to the features extracted. The time delta functions ΔA_T as given below

$$\Delta A_T(\lambda, \phi) = A_F(\lambda, \phi) - A_F(\lambda, \phi - 1), \; where \; \phi = 2, \ldots, T \tag{3}$$

The overall feature vector $A(\lambda, \phi)$ including the delta functions can be defined as:

$$A(\lambda, \phi) = [A_F(\lambda, \phi), \Delta A_T(\lambda, \phi), \Delta A_S(\lambda, \phi)] \tag{4}$$

Hence, we have extracted the features from a large speech signal corpus using AMS feature extraction.

D. Generation of Ideal Weight

The idea of generation of ideal weight has been referred from [36] to improve the speech signal quality. In this unit, the each of the individual TF units are classified into various classes by comparing with the original signal later an ideal window is found by the use of cuckoo search

Classification

Here, the input TF unit is classified into the respective class with the use of original signal and noisy signal. The classification of the speech signal to different classes is based on the Quality Ratio which is the ratio of the estimated speech magnitude $\overline{M}$ to the true speech magnitude T for each T-F unit. Here the spectrum at time slot ϕ and sub-band λ is considered; hence the quality ratio R_Q can be defined by:

$$R_Q = \frac{|\overline{M}(\lambda, \phi)|}{|T(\lambda, \phi)|} \tag{5}$$

Where estimated signal spectrum $\overline{M}$ is obtained by the product of spectrum M with the gain function G_A which is shown in below equation:

$$\overline{M}(\lambda, \phi) = G(\lambda, \phi). |M(\lambda, \phi)| \tag{6}$$

Where Gain can be found out from the equation:

$$G_A(\lambda, \phi) = \sqrt{\frac{\psi(\lambda, \phi)}{1 + \psi(\lambda, \phi)}} \tag{7}$$

Generation of Best Weight by Cuckoo Search:

Here the ideal weight of the window is generated for each of the classes making use of the cuckoo search algorithm

Step 1: Initial population

Let the noisy speech input signal be represented by M, which is defined by $M = \{m_1, m_2, \ldots, m_{Ns}\}$, where Ns is the total number of signals in the input signal. The input signal is classified into class Cl_1, Cl_2 *or* Cl_3 with the use of quality ratio. In order to obtain the best ideal binary window with less iteration, first classify the units into different classes and generate the initial window with the help classification unit. Then, fitness (SNR) is computed for the initial population to find whether it's fixed to synthesis speech enhance signal.

Step 2: New solutions

Then, with the help of initial window, generate the new window based on the equation of cuckoo search. Levi flight is performed on Y_i (initial window) to yield to get a new cuckoo Y_i^*. Considering the signal y_{i1} in Y_i, then the changed value (new solution) y_{i1}^* is given by

$$y_{i1}^* = y_{i1}^{(t+1)} = y_{i1}^{(t)} + \Lambda \otimes Levy(x). \tag{8}$$

The Levi flight equation represents the stochastic equation for random walk as it depends on the current position and the transition probability (second term in the equation). Here, the levy distribution is given by,

International Journal of Engineering Innovation & Research
Volume 2, Issue 6, ISSN: 2277 – 5668

$$Levy(x) = \sqrt{\frac{c}{2\pi}} \cdot \frac{e^{-\frac{1}{2}(\frac{c}{x})}}{x^{3/2}} \quad (9)$$

Where c is arbitrary constant. Hence, by performing Levi search, we obtain new solutions and then the fitness value (SNR value) of the new solution is found out. Let the fitness of the Levi performed nest be F_i. Initially Levi flight performed, corresponding fitness is found out F_i, compared to fitness of some other nest F_j and the replacement is carried if the condition $F_i > F_j$ is satisfied

Step 3: Termination

After the comparison and replacements, we have to abandon a fraction of worst nests and built new nests in place of them. This is done by finding the quality of the all the current nests and analyzing it. That is, keeping the best solutions and replacing the worst nests by newly built nests. Subsequently the solutions are ranked and the current best is found out. The full loop is continued till some stop criteria is met and the current best in the last loop performed will be the best solution. The ideal window weight for the training signals will be the fitness function value obtained for the best solution.

E. Waveform Synthesis Unit

In the enhancement unit, the test noisy speech signal is multiplied by the corresponding ideal binary window obtained from the cuckoo search in the training unit. Subsequently the resultant signals are synthesized to produce the enhanced speech waveforms. Here, initially the noisy speech signal is multiplied with the ideal window generated from cuckoo search algorithm directly. Suppose, Let $T(k,t)$ be noisy speech signal given as input for speech enhancement and $O(k,t)$ be the ideal window generated, Then, the enhanced signal of $E(k,t)$ is generated after applying the following equation.

$$E(k,t) = O(k,t) * T(k,t) \quad (10)$$

So, finally the original speech signal is estimated after summing the weighted responses of the 25 signal components. The spectrogram of the estimated speech signal using ideal window generation shows the level of energy similar to the original speech signal energy level at the corresponding frequencies.

IV. Results and Discussion

The results obtained and discussions based on that are given in this section. The experimental set up and simulation results are discussed in detail below. The database is utilized for getting noise sounds and it includes suburban train noise, babble, car, exhibition hall, restaurant, street, and airport and train-station noise. The sentences were originally sampled at 25 kHz and down sampled to 8 kHz and noise part is artificially added to the speech signal.

A. Experimental Set Up and Simulation Results

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 2011. The input signal is given in fig. 2, noisy signal in fig. 3 and the de-noised signal in fig. 4. The signal power is plotted for a frequency range between 0 to 2.5 KHz.

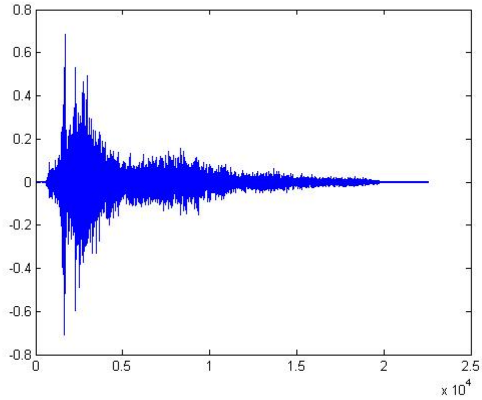

Fig.2. Input signal

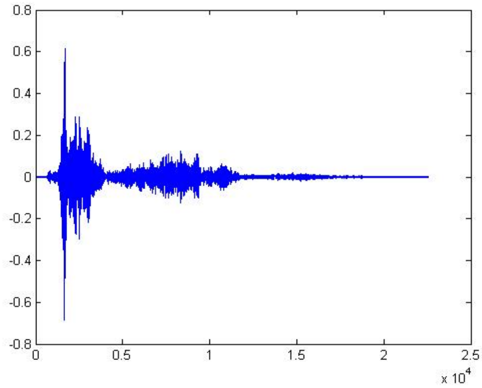

Fig.3. Noisy Signal

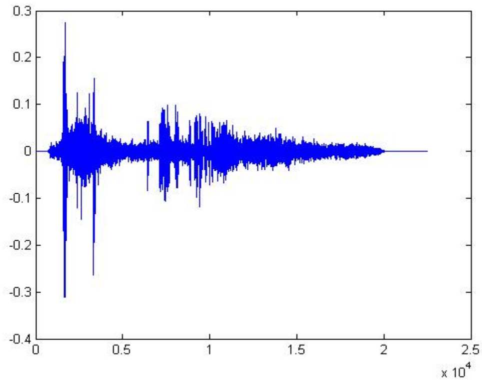

Fig.4. Denoised Signal

Table 1: Evaluation metrics for Babble Noise

Amount of Noise added	*SNR*
-5 dB	10.5691
-10 dB	17.7956
-15 dB	24.7640

International Journal of Engineering Innovation & Research
Volume 2, Issue 6, ISSN: 2277 – 5668

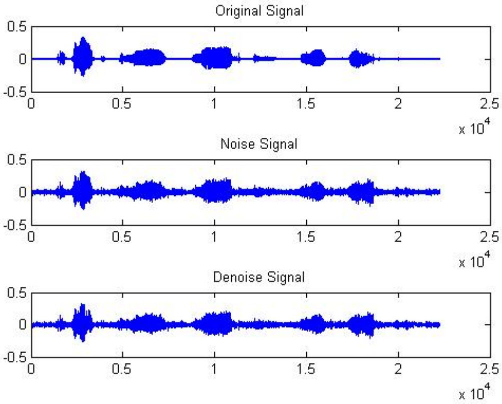

Fig.5. Input signal, Noise signal and Denoised signal for 5db Noise

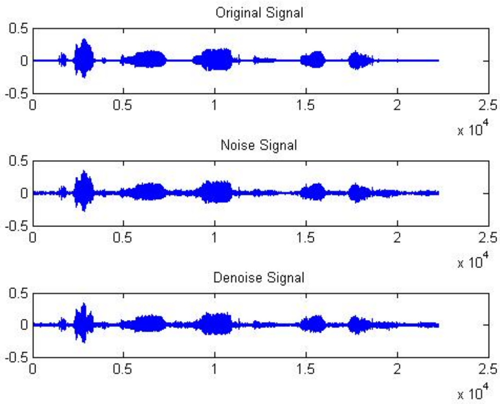

Fig.6. Input signal, Noise signal and Denoised signal for 10db Noise

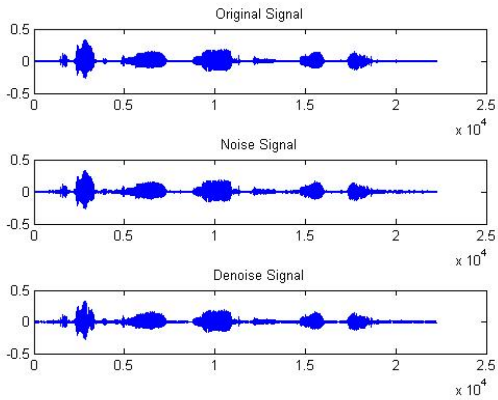

Fig.7. Input signal, Noise signal and Denoised signal for 15db Noise

Table 2: Evaluation metrics for Car Noise

Amount of Noise added	*SNR*
-5 dB	10.7005
-10 dB	17.9095
-15 dB	24.7302

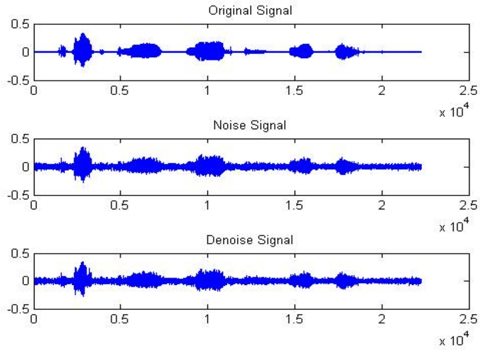

Fig.8. Input signal, Noise signal and Denoised signal for 5db Noise

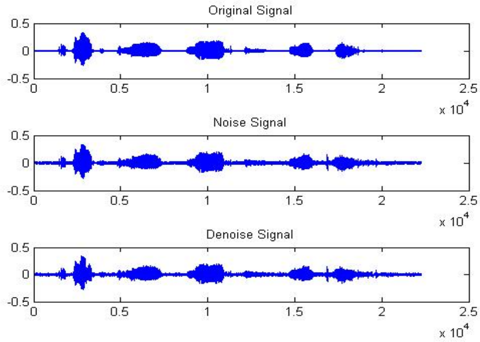

Fig.9. Input signal, Noise signal and Denoised signal for 10db Noise

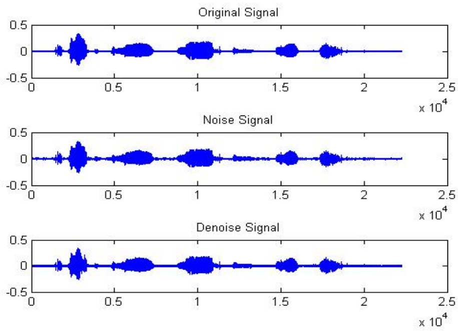

Fig.10. Input signal, Noise signal and Denoised signal for 15db Noise

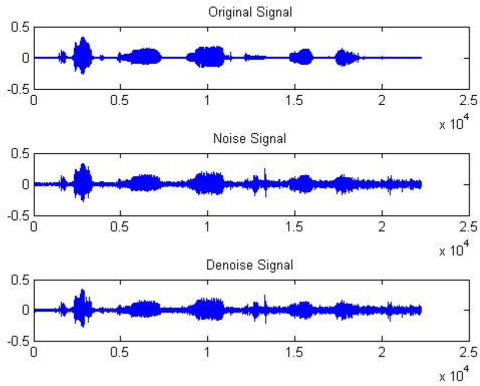

Fig.11. Input signal, Noise signal and Denoised signal for 5db Noise

International Journal of Engineering Innovation & Research
Volume 2, Issue 6, ISSN: 2277 – 5668

Table 3: Evaluation metrics for street Noise

Amount of Noise added	*SNR*
-5 dB	10.6222
-10 dB	17.6489
-15 dB	24.5211

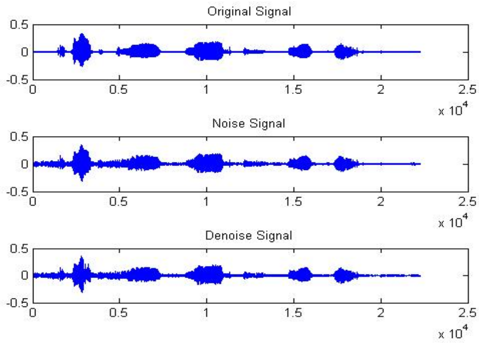

Fig.12. Input signal, Noise signal and Denoised signal for 10db Noise

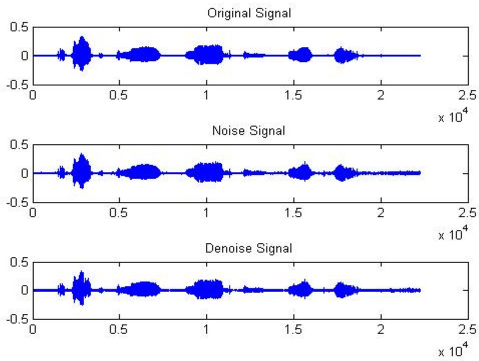

Fig.13. Input signal, Noise signal and Denoised signal for 15db Noise

Table 4: Evaluation metrics for restaurant Noise

Amount of Noise added	*SNR*
-5 dB	10.7593
-10 dB	17.7414
-15 dB	24.9270

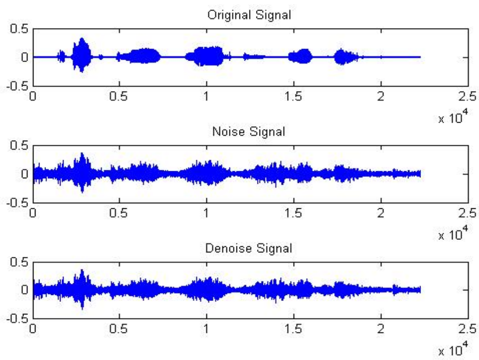

Fig.14. Input signal, Noise signal and Denoised signal for 5db Noise

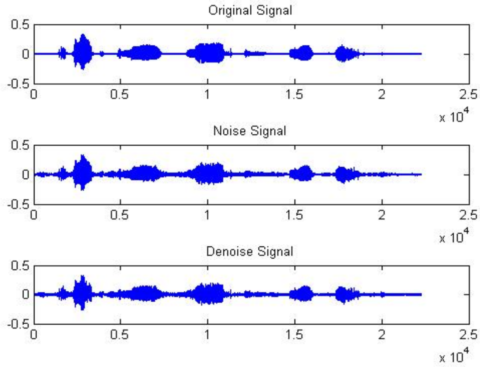

Fig.15. Input signal, Noise signal and Denoised signal for 10db Noise

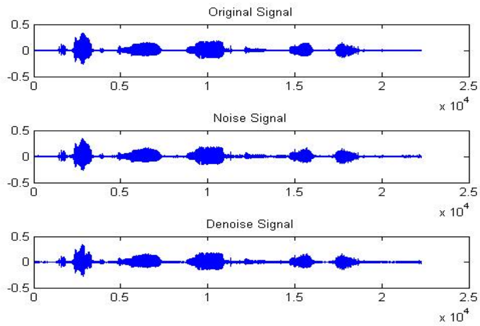

Fig.16. Input signal, Noise signal and Denoised signal for 15db Noise

Table 5: Evaluation metrics for exhibition Noise

Amount of Noise added	*SNR*
-5 dB	10.5710
-10 dB	17.8248
-15 dB	24.9196

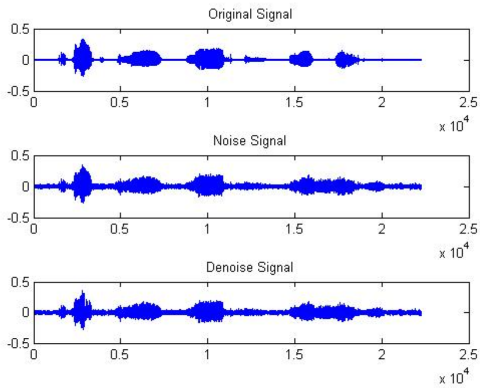

Fig.17. Input signal, Noise signal and Denoised signal for 5db Noise

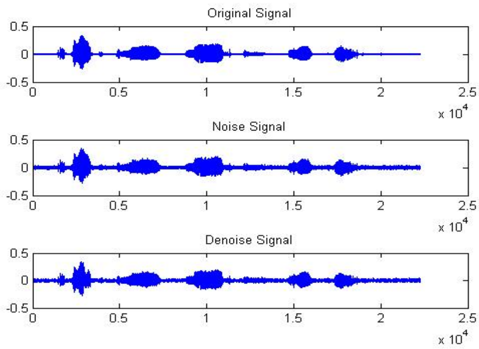

Fig.18. Input signal, Noise signal and Denoised signal for 10db Noise

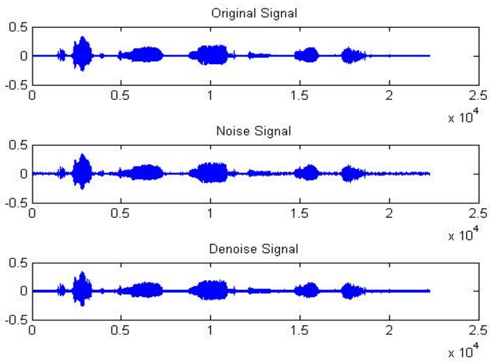

Fig.19. Input signal, Noise signal and Denoised signal for 15db Noise

From the analysis of tables (1-5) and figures (5-19) above we proved that

- In all cases, we represented Input signal Noise signal and Denoised signal for our proposed technique using the evaluation metrics
- Babble noise, train noise, car noise, exhibition noise, restaurant noise and street noise are various types of noises taken for analysis, taken at noise levels of -5dB, -10 dB and -15 dB**.**
- The evaluation metrics of SNR have been used.
- We can also see that our technique is having better average for all noise conditions.
- The results obtained proved the effectiveness of the proposed technique and is able to suppress noise and enhance the speech signal.

V. Conclusion

In this paper, cuckoo search based ideal window generation for noise suppression and enhancement of speech signal is presented. The database noises include suburban train noise, babble, car, exhibition hall, restaurant, street, and airport and train-station noise. The technique has three units: Feature extraction unit, ideal window generation unit and the waveform synthesis unit. Feature extraction is carried out using AMS and classification of signals is done to generate the initial population of cuckoo search algorithm. The Simulation of the proposed technique was carried out using various datasets. The results obtained proved the effectiveness of the proposed technique and was able to suppress noise and enhance the speech signal. Large gains in intelligibility were achieved with the proposed approach using a limited amount of training data. Overall, the summary of finding using proposed approach suggests that speech intelligibility can be improved by estimating the signal-to-noise ratio in each time-frequency unit.

References

[1] Jamal Ghasemi and Mohammad Reze Karami Mollaei, "A New Approach for Speech Enhancement Based On Eigenvalue Spectral Subtraction", *Signal Processing: An International Journal*, Vol. 3, no. 4, 2009

[2] P. C. Loizou, “Speech processing in vocoder-centric cochlear implants,” in Cochlear and Brainstem Implants, A. R. Møller, Ed. Basel, Switzerland: Advances in Oto- Rhino-Laryngology, Karger, 2006, vol. 64, pp. 109–143.

[3] P. Scalart and J. V. Filho, “Speech enhancement based on a priori signal to noise estimation,” in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2., IEEE, 1996, pp. 629–632.

[4] D.L. Donoho, “Denoising by soft thresholding,” IEEE Transactions Information Theory, 41(3), 613:627, 1995

[5] O. O. Khalifa, Z. H. Ahmad, A-H A. Hashim and T. S. Gunawan “SMaTalk: Standard Malay Text to Speech Talk System”, vol. 2, no. 5, pp. 1-16, 2008

[6] P. J. Wolfe and S. J. Godsill, “Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement,” EURASIP Journal on Applied Signal Processing, vol. 2003, no. 10, pp. 1043–1051, 2003.

[7] M. Banbrook, S. McLaughlin, and I. Mann, “Speech characterization and synthesis by nonlinear methods”, IEEE Transactions on Speech and Audio Processing, no. 7, pp. 1-17, 1999

[8] Y. Ephraim, D. Malah,: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Transactions Acoust. Speech Signal Process. ASSP, vol. 32, no. 6, pp. 1109-1121. 1984.

[9] J. R. Deller, J. H. L. Hansen, J. G. Proakis, “Discrete Time Processing of Speech Signals”, second ed. IEEE Press, New York , 2000

[10] I. Y. Soon, S. N. Koh, and C. K. Yeo, “Noisy speech enhancement using discrete cosine transform,” Speech Communication, vol. 24, no. 3, pp. 249–257, 1998.

[11] Van den Bogaert T, Doclo S, Wouters J and Moonen M, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids", J Acoust Soc Am., vol. 125, no. 1, pp. 360-371, 2009.

[12] Welker, D. P., Greenberg, J. E., Desloge, J. G., and M., Z. P. “Microphone-array hearing aids with binaural output-part II: A twomicrophone adaptive system,” IEEE Trans. Speech Audio Process, no. 5, pp. 543–551, 1997

[13] Ernst Warsitz, Alexander Krueger and Reinhold Haeb-Umbach, "speech enhancement with a new generalized eigenvector blocking matrix for application in a generalized sidelobe canceller", IEEE International Conference on Acoustics,Speech and Signal Processing, pp. 73-76,2008.

International Journal of Engineering Innovation & Research
Volume 2, Issue 6, ISSN: 2277 – 5668

[14] Y. Ephraim and D. Malah, “Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator,” IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-32, no. 6, pp. 1109–1121, 1984

[15] Lollmann, H.W. and Vary, P, "A blind speech enhancement algorithm for the suppression of late reverberation and noise", IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3989-3992, 2009.

[16] Y. Hu and P. Loizou, “Subjective comparison of speech enhancement algorithms,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 1, pp. 153–156, 2006.

[17] Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement", IEEE transactions on audio, speech and language processing, vol. 16, no. 1, 2008.

[18] P.C. Loizou, “Speech enhancement: Theory and practice,” CRC Press, 2007.

[19] Achintya Kundu, Saikat Chatterjee, A. Sreenivasa Murthy and T.V. Sreenivas, "GMM Based Bayesian Approach To Speech Enhancement In Signal / Transform Domain", In 33rd IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.

[20] A. Rezayee and S. Gazor, “An adaptive KLT approach for speech enhancement,” IEEE Trans. Speech and Audio Processing, vol. 9, no. 2, pp. 87-95, 2001

[21] Y. Ephraim, “A Bayesian estimation approach for speech enhancement using hidden Markov models,” IEEE Trans. Signal Processing, vol. 40, no. 4, pp. 725 - 735, 1992.

[22] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean square error short-time spectral amplitude estimator,” IEEE Trans. Acoust., Speech and Signal Processing, vol. ASSP-32, pp. 1109-1121, 1984

[23] Y. Ephraim and H.L. Van Trees, “A signal subspace approach for speech enhancement,” IEEE Trans. Speech and Audio Processing, vol. 3, no. 4, pp. 251-266, 1995

[24] S. Quackenbush, T. Barnwell, and M. Clements, Objective Measures of Speech Quality. Englewood Cliffs, NJ: Prentice-Hall, 1988

[25] L. Thorpe and W. Yang, “Performance of current perceptual objective speech quality measures,” in Proc. IEEE Speech Coding Workshop, pp. 144–146, 1999.

[26] Y. Lu and P.Loizou, “Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty,” IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no. 5, pp. 1123–1137, 2011

[27] Youyi Lu., Martin Cooke,”The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise”, Speech comm.(51), pp. 1253-1262, 2009

[28] D. Brungart, P. Chang, B. Simpson, and D. Wang, “Isolating the energetic component of speech-on-speech windowing with ideal time-frequency segregation,” J. Acoust. Soc. Amer., vol. 120, pp. 4007–4018, 2006

[29] N. Li and P. C. Loizou, “Factors influencing intelligibility of ideal binary- windowed speech: Implications for noise reduction,” J. Acoust. Soc.Amer., vol. 123, no. 3, pp. 1673–1682, 2008

[30] Xin-She Yang,"Cuckoo Search via Lévy flights", Nature & Biologically Inspired Computing, 2009.

[31] Brady Laska, Miodrag Bolic, Rafik Goubran, "Discrete cosine transform particle filter speech enhancement", Speech Communication, vol. 52, pp. 762–775, 2010.

[32] Brady N. M. Laska, Miodrag Bolic, and Rafik A. Goubran, "Particle Filter Enhancement of Speech Spectral Amplitudes", IEEE transactions on audio, speech, and language processing, Vol. 18, No. 8, November 2010.

[33] Gibak Kim, Yang Lu, Yi Hu, and Philipos C. Loizoua, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners", J. Acoust. Soc. Am., vol. 126, no.3, 2009.

[34] Gibak Kim and Philipos C.Loizou, "Improving Speech Intelligibility in Noise Using Environment Optimized Algorithms", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 8, November 2010

[35] Gibak Kim and Philipos C. Loizou, "Reasons Why speech- enhancement algorithms do not improve speech intelligibility and suggested solutions", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 1, January 2011

[36] Anil Garg and 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 (In Press)

AUTHOR’S PROFILE

Anoop.V
Research Scholar, Department of Electronics & Communication Engineering, T-John Institute of Technology, Bangalore-India

Dr. P. V. Rao
Prof. & R&D Head, Department of Electronics & Communication Engineering, T-John Institute of Technology, Bangalore-India