ABSTRACT
Speech signals from the uninhibited environment is the combination of degraded and required speech components. The degraded noise can reduce the quality and intelligibility of the speech signal considerably. The degraded speech therefore wants to be processed for the improvement of speech components. Improvement of speech quality is done by speech enhancement methods where various algorithm techniques are used and hence by this the background noise that mainly causes interference can be reduced considerably. In this paper a new speech enhancement technique is proposed to single channel speech enhancement systems on recently developed algorithm called Firefly algorithm. FA is a nature inspired, metaheuristic and stochastic algorithm which uses a kind of randomization on searching for a set of solutions. FA has some important behaviors to be considered such as attractiveness and light intensity variation based on this solution will be found. This paper aims to present the firefly algorithm as a very potential optimization technique in the context of speech enhancement, where the proposed technique is added with the cosh window in the preprocessing module and the multiblade filtering technique in the spectral filtering module. Simulation results of FA is observed to be 37.15 dB SNR and PESQ is 1.32 out of 4.5 even at 0dB babble noise and hence this can be used in speech enhancement systems for the improvement in terms quality.

KEYWORDS: Single channel speech enhancement, Metaheuristic algorithms, stochastic, Firefly algorithm, Cosh window, Multi blade filter.

INTRODUCTION
Speech enhancement algorithms attempt to improve the performance of communication systems when their input or output signals are corrupted by noise. The presence of background noise causes the quality and intelligibility of speech to degrade. Here, the quality of speech refers how a speaker conveys an utterance and includes such attributes like naturalness and speaker recognisability. Intelligibility is concerned with what the speaker had said, that is, the meaning or information content behind the words. Therefore, a noisy environment reduces the speaker and listeners ability to communicate. To reduce the impact of this problem speech enhancement can be performed. There are many ways to classify speech enhancement methods. It is usually difficult for a typical algorithm to be able to perform homogeneously across all noise types. Therefore, usually a speech enhancement system is based on certain assumptions and constraints that are typically dependent on the application and environment. The speech enhancement systems can be classified based on number of input channels (one/two/multiple), domain of processing (time/frequency) and type of algorithm (Non adaptive/Adaptive). There are lots of literature works regarding speech enhancement methods. The Particle Swarm Optimization Algorithm and MMSE filter technique were compared with MMSE and BNMF for effective enhancement[1]. Bat algorithm is analyzed with the standard PSO[2] where the BA excels in the obtained results. Bat Algorithm (BA) and Particle Swarm Optimization Algorithm (PSO) comparison where the algorithms are...
trained to Radial Basis function network (RBF) [3] to classify the efficient algorithm with the benchmarking dataset. A Comprehensive review of Firefly algorithm is made which will be our proposed technique [4]. Three types of metaheuristic algorithms called firefly algorithm, bat algorithm and cuckoo search algorithm were compared to find optimal solutions and series of computational experiments using each algorithm were conducted and after the analysis it is observed that firefly algorithm seems to perform better than bat algorithm and cuckoo search.[5] A comparative study of different window functions such as Hanning, Hamming, Blackman, Cosh, kaiser and Exponential windows and other transforms such as discrete Fourier transform (DFT), discrete cosine transform (DCT), discrete wavelet transform (DWT), Ridgelet transform (RT) has been made[6][7]. Hybrid Wiener spectrogram filter (HWSF) has been made[8]. The existing approaches are PSO and BA algorithm, where the novel approaches have their disadvantage such as complexity, functional difficulty and slow convergence speed. Hence the present study focuses to overcome the disadvantages in the existing algorithms with PSO and BA which use the MMSE filtering technique in the post processing module. The metaheuristic, nature inspired, stochastic Firefly algorithm overcomes the drawback of PSO and BA algorithm. FA algorithm is a metaheuristic stochastic optimization algorithm which is based on the light intensity and attractiveness of the fireflies. The FA algorithm has a property of randomness in searching a set of solutions. This algorithm is used with the cosh window in the pre-processing module and multiblade filtering technique in post processing module.

In this paper, the Firefly algorithm is used to find the effectiveness in speech enhancement. The rest of the paper is organised as follows. Section II describes the proposed speech enhancement algorithm. Section III, describes the objective measures used for the evaluation results of algorithm. Section IV conclusions are given finally.

Methodology Used:

In this study pre-processing module, optimization module and spectral filtering modules are carried out and the block diagram is given for this proposed technique in Fig. 1.

![Block Diagram of Speech Enhancement System](image)

Fig. 1: Block Diagram of Speech Enhancement System.

A. Pre-Processing Module:

Framing:

Initially, framing is carried out where the input signals are divided into frames to achieve stationarity. The process of division of signals into frames are carried out and the resultant frames of 20ms correspond to n samples.

\[
n = t_{ed}
\]

The Speech is divided into four frames. Each frames first part will be shared with the previous frame and the last part with the next frame. The time frame step \(t_{fs}\) indicates the start time of each frame. The overlap starts at the time where a new frame starts and the current frame stops.

\[
t_{fl} = t_{fs} + t_{0}
\]

where, In Eqn. 2, frame length \(t_{fl}\) is given.

Cosh Window:

The shape of a window in time domain decides the characteristics of resultant filter in frequency domain. Several window functions are available in literature. For the present work, we have selected three parameter cosh window for truncation of infinite impulse response. Cosh window can be defined as

\[
w[n] = \frac{\cosh(\alpha(\frac{2n}{T} - \frac{1}{2}))}{\cosh(\alpha)}; 0 \leq n \leq N-1;
\]

The hyperbolic cosine of \(y\) is defined in Eqn. 4,

\[
Cosh (q) = (e^q + e^{-q})/2
\]

The proposed window has been derived in the similar way as that of Kaiser Window, but it has computational cost advantage since because it has no power series expansion in its time domain representation. On obtaining the spectrum design equations for the proposed system function, it is compared with the Kaiser and ultraspherical windows in terms of various spectral characteristics. Simulation results show that the proposed window
performs better on consideration of roll-off ratio of the side lobe compared to Kaiser window for the same window length and normalized main lobe width. This window provides better ripple ratio as well as side lobe roll-off ratio when compared with the ultra-spherical window. Performance of this system seems to be better when it is combined with the Hamming window in terms of the ripple ratio, than the performances of Kaiser and Hamming window combination. Moreover, the paper presents the relevance of the proposed window in FIR filter design. The result shows that the filters designed using cosh window provide better stop band attenuation than the filters designed using the Kaiser window.

Fast Fourier Transform:
Decomposition of N point time domain signal into N point frequency domain signals and each composed of a single point. Secondly, N frequency spectra calculation corresponding to the time domain signals. Lastly, synthesized into a single frequency spectrum.

B. Optimization Module:
Firefly Algorithm:
Firefly are unisex so that one fire fly will be attracted to another regardless of sex. Attractiveness is proportional to the brightness and hence the less brighter one will move towards the other. Attractiveness decreases as their distance increases and if there is no brighter one than a particular firefly. There are two important issues
• variation of light intensity and
• formulation of attractiveness
For simplicity we can always assume that attractiveness of a firefly is determined by its brightness which in turn associated with the objective function.

Light Intensity and Attractiveness:
For optimization problems, the brightness I of the firefly at location X can be equated as
\[ I(X) \propto f(X) \quad (5) \]
The attractiveness \( \beta \) is relative and is determined by the following Eqn. 6.
\[ \beta(x) = \beta_0 e^{-\gamma x^2} \quad (6) \]
where \( \beta_0 \) is the attractiveness at the position \( x=0 \). In Eqn. 7, distance \( r_{ij} \) between any two fireflies \( i \) and \( j \) with \( x_i \) and \( x_j \) respectively.
\[ r_{ij} = || x_i - x_j || \quad (7) \]
Light intensity decreases with the distances from the source and the light is also absorbed by the media. The attractiveness varies with the degree of absorption. The global maxima and minima can be found simultaneously. In a further improvement on the convergence speed of the algorithm, randomization parameter \( \alpha \) can be varied. Movement of firefly \( i \) that is located in place \( x_i \) toward brighter firefly \( j \) that is located in place \( x_j \) is shown by the Eqn. 8,
\[ x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma r^2} (x_j - x_i) \quad (8) \]

C. Spectral Filtering Module:
Multiblade Filtering:
Residual noise is removed. In this post processing step, the leftover non speech and noise components are replaced by the median values. It is observed from the spectrograms and identified through listening that the musical tones appear randomly in the spectrum as isolated peaks or short ridges. Moreover, the speech content in the spectrograms exhibit distinguishing features. These observations prompted us to use 2D spectrogram for enhancement. The spectrogram of a speech signal has distinct characteristics, and we need to focus on different directions relative to a point of interest (POI) for nonspeech components. A total of 16 blades are used to indicate the different orientations. The first 6 blades focus on different 7-by-7 orientations with the POI in the centre. The other which includes five left blades and five right blades. Along with these orientations, investigation on the properties are carried out in order to classify the POI into speech components or nonspeech components accordingly.

RESULTS AND DISCUSSION

Five clean speech sentences are selected randomly from NOIZEUS database for the evaluation of the proposed system performance. This database contains 30 IEEE sentences produced by three male and three female speakers (five sentences /speaker), and was corrupted by eight different real-world noises at different SNRs. The noise signals were added to the speech signals at SNRs of 0, 5, 10, and 15dB. From NOIZEUS database, different noise signals are added to the speech signal and are denoised using FA algorithm with different Signal to Noise Ratio (SNR) levels.
Fig. 2: Waveform of Input Noisy Speech.

Fig. 3: Spectrogram of Input Noisy Speech.

Fig. 4: Waveform of Enhanced Speech.

Fig. 5: Spectrogram of Enhanced Speech.

Fig 2 shows the Waveform of Input Noisy Speech Signal. Fig 3 shows the Spectrogram of Input Noisy Speech Signal. The spectrogram of noisy signal corrupted with babble noise of 0dB SNR. The speech signal is separated into several frames. For this 0 dB input SNR, when the corrupted signal is run through the FA algorithm, the output SNR is increased. Fig 4 shows the Waveform of Enhanced Speech Signal. The figure shows that the harmonic part is visible and hence the harmonic part of the signal is restored. Fig 5 gives the enhanced speech spectrogram of the proposed technique where the visible harmonics shows the noise reduction of the babble noise of 0dB.

A. Objective Measures:

The objective measures are made to Signal to noise ratio (SNR) and Perceptual evaluation of speech quality (PESQ) of Firefly algorithm.

B. SNR Estimation:

SNR is one of the evaluation parameter for measuring the enhancement of the algorithm. SNR is defined as the difference between the SNR of the enhanced speech and that of the noisy speech and is calculated by

$$\frac{S}{N} = 10 \log_{10} \left[ \frac{\sum x^2(n)}{\sum (x(n) - y(n))^2} \right]$$

where, In Eqn. 9, x(n) is the clean speech and y(n) is the distorted speech. Different noises at various levels can be obtained.
C. PESQ Estimation:
Computation of PESQ is done by equalising the clean signal and the degraded signal to a standard listening level and then passed through the filter. Symmetric disturbance which is the measure of absolute audible error. Asymmetric or additive disturbance which are the measure of the audible errors that are louder than the reference.

\[ PESQ = 4.5 - a_1 D_{adj} - a_2 A_{adj} \]

where \( a_1 = -0.1 \) and \( a_2 = -0.0309 \)

Fig 10 shows the representation of PESQ (out of 4.5) for different input noises level in the firefly algorithm.

| Table 1: Results of output SNR with different Input SNR. |
|-------------------------|-------------------------|
| NOISE TYPE              | INPUT SNR(dB) | OUTPUT SNR(dB) |
| BABBLE NOISE            | 0 dB          | 37.15          |
|                         | 5 dB          | 32.50          |
|                         | 10 dB         | 32.82          |
|                         | 15 dB         | 31.79          |
| AIRPORT NOISE           | 0 dB          | 33.48          |
|                         | 5 dB          | 33.32          |
|                         | 10 dB         | 32.38          |
|                         | 15 dB         | 32.97          |
| CAR NOISE               | 0 dB          | 35.13          |
|                         | 5 dB          | 31.41          |
|                         | 10 dB         | 33.52          |
|                         | 15 dB         | 33.75          |

Table 1 shows the output SNR values of babble noise, airport noise and car noise of 0, 5, 10 and 15dB it can be seen that the output SNR level is 37.15dB of babble noise and the quality in speech is achieved.

**Fig. 6:** Output SNR level for different Input SNR.

Fig 6 shows the SNR for different input SNR such as airport, babble, car, noises at 0, 5, 10 and 15dB in the Firefly algorithm.

| Table 2: Results of PESQ with different Input SNR. |
|-------------------------|-------------------------|
| NOISE TYPE              | INPUT SNR(dB) | OUTPUT PESQ(dB) |
| BABBLE NOISE            | 0 dB          | 1.32           |
|                         | 5 dB          | 1.26           |
|                         | 10 dB         | 1.30           |
|                         | 15 dB         | 1.36           |
| AIRPORT NOISE           | 0 dB          | 1.45           |
|                         | 5 dB          | 1.37           |
|                         | 10 dB         | 1.35           |
|                         | 15 dB         | 1.40           |
| CAR NOISE               | 0 dB          | 1.51           |
|                         | 5 dB          | 1.44           |
|                         | 10 dB         | 1.37           |
|                         | 15 dB         | 1.39           |

Table 2 shows the output PESQ values of different noise levels such as babble noise, airport noise and car noise of 0, 5, 10 and 15dB it can be seen that the output PESQ level is 1.51 out of 4.5 of car noise which is reliably good.
Fig. 7: Different Input noise levels vs PESQ.

Fig 7 shows the representation of PESQ (out of 4.5) for different input noises levels such as airport, babble and car noise at 0, 5, 10 and 15 dB in Firefly algorithm.

Conclusion:
The Firefly algorithm offers a new direction for speech enhancement as it exhibits many advantages such as high convergence rate and minimum computation. Complexity and difficulty level of functions had no effect to FA. FA also proves better in terms of speed of convergence which is due to the effect of generating completely different random numbers to be used in iterative procedures of algorithm. FA seems to be potentially more powerful and favorable optimization tool due to effect of attractiveness function which is unique to firefly behavior FA performs local search well as in some algorithms, it is unable to completely get rid of local search. FA remembers the history of better solution so as to achieve the high optimization solution.

REFERENCES