SINGLE CHANNEL SPEECH ENHANCEMENT USING EVOLUTIONARY ALGORITHM WITH LOG-MMSE

Kalpana Ghorpade[a*], Arti Khaparde[b]

[a]E &TC department, Faculty of Engineering, MKSSS's Cummins College of Engineering for Women, Pune, Maharashtra, India
[b]Department of ECE, Faculty of Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India
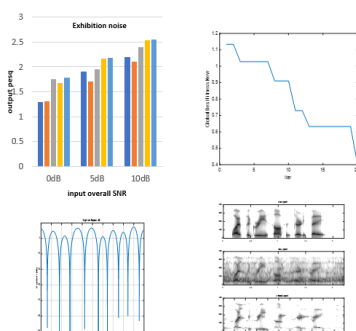
*Corresponding author
kalpana.joshi@cumminscollege.in

Graphical abstract

Filter with multiple passbands and notches for speech enhancement using modified APSO. The modified algorithm converges faster giving optimum filter coefficients and improvement in PESQ value and Segmental SNR.

Abstract

Additive noise degrades speech quality and intelligibility. Speech enhancement reduces this noise to make speech more pleasant and intelligible. It plays a significant role in speech recognition or speech-operated systems. In this paper, we propose a single-channel speech enhancement method in which the log-minimum mean square error method (log-MMSE) and modified accelerated particle swarm optimization algorithm are used to design a filter for improving the quality and intelligibility of noisy speech. Accelerated particle swarm optimization (APSO) algorithm is modified in which a single dimension of particle position is changed in a single iteration while obtaining the particle's new position. Using this algorithm, a filter is designed with multiple passbands and notches for speech enhancement. The modified algorithm converges faster compared with standard particle swarm optimization algorithm (PSO) and APSO giving optimum filter coefficients. The designed filter is used to enhance the speech. The proposed speech enhancement method improves the perceptual estimation of speech quality (PESQ) by 17.05% for 5dB babble noise, 33.92 % for 5dB car noise, 14.96 % for 5dB airport noise, and 39.13 % for 5dB exhibition noise. The average output PESQ for these four types of noise is improved compared to conventional methods of speech enhancement. There is an average of 7.58 dB improvement in segmental SNR for these noise types. The proposed method improves speech intelligibility with minimum speech distortion.

Keywords: Evolutionary algorithms, Log-MMSE, Particle swarm optimization, Speech enhancement, Speech intelligibility

1.0 INTRODUCTION

Speech signals get contaminated by background noise making them less intelligible. In speech-operated systems, degraded speech affects the performance of the system. Noise generated by vehicles, co-speakers, street noise gets added to the speech signal. In a real environment, complete noise cancellation is not possible as it is difficult to track varying noise types and characteristics that change with time[1]. But reducing additive noise in speech is possible to make speech more intelligible and to enhance the efficiency of speech-dependent applications. This is done by speech enhancement systems. Speech enhancement has gained a lot of research interest. For single-channel speech enhancement systems, enhancing speech is quite challenging as

there is no reference to noise. A variety of speech enhancement algorithms are available which improve the quality of speech [2,3]. Spectral subtraction for speech denoising is suggested by Boll [4,5]. The output of the spectral subtraction method suffers from musical noise. To overcome this problem, the modulation frequency domain is suggested as the transform of time series of acoustic frequency [6]. Later, it is proposed that the first short time Fourier transform (STFT) of the time domain noisy speech as the acoustic spectrum and the second STFT of the time series at a particular acoustic frequency as the modulation spectrum at that frequency [7]. The strength of the modulation domain as compared to the acoustic domain is evaluated for spectral subtraction in [8]. It is concluded that musical noise gets removed when spectral subtraction is carried out in the

modulation domain [8,9]. In [10], modulation domain Kalman filter and subspace method are implemented for speech enhancement. In [11], authors have implemented single-channel speech enhancement by use of Kalman filter in log power spectral domain in which Bayesian estimate of speech and noise is used in the update step. In [12], Gaussring model-based modulation domain Kalman filtering is proposed in which noise and speech models are estimated separately and combined in modulation domain Kalman filter to give an enhanced speech.

For noise cancellation in speech, adaptive filter algorithms like least mean square (LMS), normalized least mean square (NLMS), recursive least square (RLS) algorithms which are based on gradient descent, are used to design adaptive filters [13,14,15]. For multimodal error surfaces, these algorithms are not suitable as there is a problem of local minima giving sub-optimal solutions [16]. Evolutionary algorithms can give a satisfactory solution to multimodal problems. These optimization algorithms are independent of system structure which use objective function information. They can take care of optimality on rough, discontinuous, and multimodal surfaces [17]. Particles of particle swarm optimization (PSO) make use of current knowledge about the search space while deciding the search area in the next iteration. This exploitative nature of PSO is useful in finding a potential solution to the problem but sometimes it leads to premature convergence at local optima [18]. This is especially true for gbest versions of PSO, in which all particles move towards the current global best solution. When all particles are attracted to a single position in the search space, there is less exploration of the search space. To overcome this, Eberhart and Kennedy suggested nbest PSO algorithm in which there are smaller neighborhoods to each particle. It is less prone to premature convergence than the gbest algorithm. Still, it suffers from convergence to the local optima. So different variants of standard PSO are suggested in the literature to improve efficiency.

In [18], modifications are done in the basic PSO to improve efficiency and speed of search, to have modified PSO (MPSO) which is used for adaptive infinite impulse response (IIR) structure. In [19], the author suggested accelerated PSO (APSO) to accelerate the speed of convergence in which the velocity vector uses only the global best. Various applications of APSO are discussed in [20]. In [21], craziness-based PSO (CRPSO) is used to design finite impulse response (FIR) high pass filter. The global search ability of basic PSO is improved in CRPSO by modifying the velocity equation. It is used in [22] to have CRPSO based design of a band stop filter. FIR high pass filter and bandstop filter are designed in [23] by using evolutionary algorithms like PSO, real coded genetic algorithm (RCGA) and cuckoo search algorithm (CSA). The magnitude response of these filters is compared with the magnitude response of filters designed by the Parks McClellan (PM) algorithm. In [24], the improved learning strategy is suggested which replaced the social and cognitive factors of PSO and guided the search direction. The use of constriction factor along with optimal weights for social and cognitive factors is proposed in optimal PSO in [25]. In [26], a novel dimensional learning mechanism is put forth to overcome the shortcoming of PSO. In [27], new regions of search space are found to improve the search for the global best particle. The review of different variants of PSO along with their applications is done in [28].

PSO and its variants are implemented for speech enhancement applications. In [29], APSO is used for dual channel speech enhancement and the results are compared with simple PSO. A combination of PSO and gravitational search algorithm (hybrid PSOGSA) is suggested in [30] for dual-channel speech enhancement which gives an improvement in output signal to noise ratio (SNR) compared to PSO and GSA. In [31], to increase diversity, a shuffled sub swarm approach is suggested for dual-channel speech enhancement application. The directed searching approach is implemented for dual-channel speech enhancement in [32]. For single-channel speech enhancement, a hybrid model of spectral filtering and PSO, a combination of minimum mean square error and PSO (MMSEPSO) is done in [33]. In [34], single-channel speech enhancement is done in which PSO is used for parameter optimization while classifying voiced and unvoiced speech. The contribution of this research is -

- Formulation of a single-channel speech enhancement system based on multi-band filter
- Modification of accelerated PSO algorithm for efficient convergence and better objective measure values.
- Use of proposed (modified) algorithm to design a multi-band filter.
- Formulation of the objective function
- Improvement in convergence speed as compared to standard evolutionary algorithms
- Improvement in the perceptual estimation of speech quality (PESQ) of enhanced speech
- Improvement in segmental SNR of enhanced speech

The proposed algorithm finds the most optimized filter coefficients for speech enhancement. We examine the suitability of the suggested algorithm by comparing its convergence speed and improvement in objective measures like the perceptual estimation of speech quality (PESQ) and segmental SNR with that of standard PSO and conventional APSO. Also, the results of the proposed speech enhancement method are compared with the results of other speech enhancement techniques.

2.0 METHODOLOGY

Standard Particle Swarm Optimization (PSO) and Accelerated PSO

Particle Swarm Optimization was introduced by Eberhart and Kennedy [35]. Here, each particle is nothing but a possible solution to the problem. It moves through the problem hyperspace [35]. Vector m_i gives the position and n_i the velocity of the ith particle in the tth iteration.

$$m_i^t = m_i^{t-1} + n_i^t \qquad (1)$$

$$n_i^t = W * n_i^{t-1} + \psi_1 * rand_1 * (p_i - m_i^{t-1}) + \psi_2 * rand_2 (p_g - m_i^{t-1}) \qquad (2)$$

Where ψ_1 and ψ_2 are the positive numbers and, are random numbers with uniform distribution in the range of [0,1]. W is inertia constant. In every iteration, the fitness function value is calculated for every particle. The best value among them (minimum or maximum depending on the objective function) gives the global best particle of that iteration giving the global

best position (p_g). Comparing every particle's current fitness value with its fitness value of the previous iteration, the local best of that particle is decided to give the local best position (p_i). The velocity of the particles is updated based on these two components and inertia as given in equation (2). Accordingly, positions are updated for the next iteration given in equation (1). The process is repeated until the stopping criteria are met giving an optimal solution to the problem [35]. To increase the diversity in the solutions, the individual best is introduced in PSO. However, this diversity can be simulated using some randomness. Accelerated particle swarm optimization (APSO) is suggested in [19,20] in which to accelerate the convergence of the algorithm, only global best is suggested. Thus, eqn. (3) gives the particle position and eqn (4) the velocity in the APSO model. Here, the velocity of particles is dependent only on the global best of the swarm. Position update eqn. (3) remains the same as eqn.(2).

$$m_i^t = m_i^{t-1} + n_i^t \tag{3}$$

$$n_i^t = n_i^{t-1} + A * rand_1 + B * rand_2 (p_g - m_i^{t-1}) \tag{4}$$

where A is drawn from N (0, 1), t relates to the current iteration, (t-1) the previous iteration, i indicates i^{th} particle. The range of values of constants A and B of APSO is 0.1 to 0.5 and 0.1 to 0.7 respectively [20].

Linear Phase FIR Notch Filter

The amplitude response of ideal multiple notch filter with notch frequencies at $\{\omega_i\}$ where

i= 1 to r is given in [36]

$$
\begin{aligned}
H_d(\omega) &= 0 \quad for \quad \omega \in \Omega^0 \\
&= 1 \quad for \quad \omega \in \Omega^1
\end{aligned}
\tag{5}
$$

Where Ω^0 and Ω^1 are the set of frequencies given by the equation (6)

$$
\begin{aligned}
\Omega^0 &= \omega \mid |\omega - \omega_i| \le \frac{\Delta\omega}{2}, 1 \le i \le r \\
\Omega^1 &= [0, \pi] - \Omega^0
\end{aligned}
\tag{6}
$$

The zero-phase frequency response of type I linear phase finite impulse response (FIR) filter with even order ($N=2M$) and symmetrical impulse response is given as

$$H_0(\omega) = h(M) + 2 \sum_{m=1}^{M} h(M-m)\cos(m\omega) \tag{7}$$

the frequency response of type I linear phase FIR filter is given as

$$H(e^{j\omega}) = H_0(\omega)e^{-jM\omega} \tag{8}$$

Proposed Filter by Single Dimension Change APSO (SDCAPSO)

Multi-channel speech enhancement methods give superior results but with extra hardware for taking the spatial information. Here, we propose a filter with multiple passbands and stopbands with the use of single dimension change APSO (SDCAPSO) for single-channel speech enhancement. In conventional APSO, particle position is estimated by adding velocity in each dimension of the position vector. So, every dimension of the position vector gets modified simultaneously. Due to simultaneous changes in all dimensions, the potential solution may be lost. Here, we have implemented a change in a single dimension of position, in a single iteration keeping the rest of the dimensions as they were in the previous iteration for all particles. By using this single dimension change APSO, we present here the optimal design of an even order linear phase-type I finite impulse response (FIR) filter with multiple passbands and stopbands and symmetric impulse response h(n) for speech enhancement. Initially, we used log-MMSE for preprocessing the noisy speech. It reduces noise without affecting the speech signal [2]. The output of log-MMSE is framed with a frame size of 25 ms and 2.5 ms overlap. It is windowed by a hamming window and applied as the input to the filter designed by a single dimension change APSO(SDCAPSO). Coefficients of filter are considered as the particles of SDCAPSO. Here we have considered symmetrical h(n). Due to symmetry in h(n), the dimension of the search space gets reduced to *(M/2) +1* where *M* is the order of the filter. The order of the filter is 24. That means there will be 25 coefficients (length of the filter) for the filter. Due to the symmetrical nature of h(n), the problem gets reduced to deciding only the first 13 coefficients by using the optimization algorithm. Thus, there will be 13 coefficients for each particle giving the dimension of the particle position as 13. Such 100 particles (population size) are initialized randomly at the start of the algorithm. The number of iterations is kept equal to the integer multiple of the number of coefficients of a particle. *nn* is the dimension of the particle which ranges from 1 to 13 here. For iteration value equal to 1 to 13, *nn = Iter* value where *Iter* is the current iteration value. But when *Iter* is more than 13, the nn value is given by Equation (10) in which *yy* is the integer set to 1 at the start of the algorithm. *yy* is incremented by 1 after every 13 iterations. The position update Eqn. is given by (9) and (10) here.

$$m_i^t(nn) = m_i^{t-1}(nn) + n_i^t(nn) \quad nn = Iter \quad if\, Iter <= 13 \tag{9}$$

$$nn = Iter - yy * 13 \quad if\, Iter > 13 \tag{10}$$

Objective Function

As said before, the particles of the algorithm are the coefficients of the filters with length *(M/2) +1*. Every single particle represents a single filter. For every particle (filter), frequency response is evaluated. The magnitude response of each such filter is divided into 16 bands each containing 16 frequency components. A filter with a magnitude response value equal to one up to *Fs/2* is considered as a reference filter. The magnitude response of the reference filter is also divided into 16 bands. For every particle, every frequency component in each frequency

band of magnitude response is compared with the corresponding frequency component of the reference filter. For the first twelve bands, the threshold value (TH) for comparison is set to 0.001 and for the remaining four bands to 0.1. In each band gain value, $G11$ is set for every frequency component as follows, and using it, the gain per band ($G1$–$G16$) is initialized as

$$
\left.
\begin{aligned}
G11 &= 0 \quad \text{if } \left(\left|H_D\left(e^{j\omega}\right)\right| - \left|H_R\left(e^{j\omega}\right)\right|\right) < \text{TH} \\
G11 &= \left(\left|H_D\left(e^{j\omega}\right)\right| - \left|H_R\left(e^{j\omega}\right)\right|\right) \quad otherwise \\
G11 &= mean\ (G11)
\end{aligned}
\right\} \quad (11)
$$

Based on the value of G11, difference D is evaluated as-

$$
\left.
\begin{aligned}
D &= \left(\left|H_D\left(e^{j\omega}\right)\right| - \left|H_R\left(e^{j\omega}\right)\right|\right) \quad if\ G11 = 0 \\
D &= G1 * \left(\left|H_D\left(e^{j\omega}\right)\right| - \left|H_R\left(e^{j\omega}\right)\right|\right) \quad otherwise
\end{aligned}
\right\} \quad (12)
$$

In this way, D for the whole frequency range (for 256 frequencies) is evaluated. The range of 256 frequencies is again divided into 2 slots.

$$
\left.
\begin{aligned}
J1 &= \max\ (D(1:232)) \\
J2 &= \max\ (D(233:256)) \\
J &= ObjFunction = J1 + J2
\end{aligned}
\right\} \quad (13)
$$

The algorithm tries to find the minimum value of the objective function by optimizing the filter coefficients.

Selection of Population Size

To select the optimum swarm (population) size, we varied the population size as $N=30$, $N=50$, $N=70$, and $N=100$ and ran the algorithm for the maximum number of iterations set in the algorithm to achieve the minimum value of the objective function. With a smaller population size, the algorithm takes a greater number of iterations for optimization of the objective function. $N=30$ and $N=50$ need a larger number of iterations compared to $N=100$ and $N=70$. A minimum number of iterations are needed by $N=100$ to give the smallest objective function value. So, we selected the swarm size of $N=100$ particles. The convergence of the algorithm for various population sizes is shown in Figure 1.
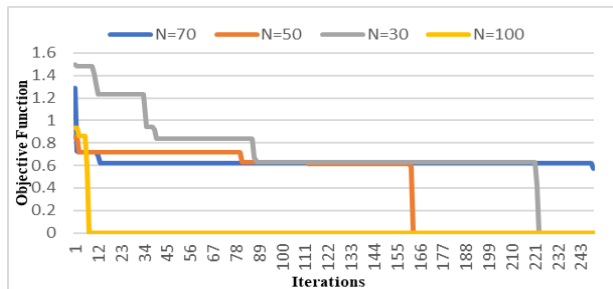
Selection of Constants A, B

The range of values of constants A and B of APSO is $A=0.1$ to 0.5 and $B=0.1$ to 0.7 [20]. Considering this range, we took different combinations of A and B and executed our algorithm. We kept the threshold value very small so that for each combination of A and B the algorithm is allowed to run through a maximum number of iterations. The convergence profile for various pair-wise values of (A, B) for the proposed algorithm (SDCAPSO) is given in Figure 2. For $A=0.3$, $B=0.5$, the algorithm converges in a smaller number of iterations (number of iterations =583) to the set threshold value ($TH = 0.2$) compared to other sets of (A, B). So, we selected the values as $A=0.3$, $B=0.5$. Table1 gives parameter values selected for the proposed algorithm (SDCAPSO).

Convergence Comparison of PSO, APSO, SDCPSO and the Proposed Algorithm

We compared the convergence result of the proposed algorithm (SDCAPSO) with that of single dimension change PSO (SDCPSO), PSO, and APSO. For comparison, we kept all the parameters of the algorithms the same along with the objective function equation and ran the algorithms. We kept the threshold value at 0.4 for the objective function so that the algorithms will go through the maximum number of iterations to achieve it. The parameters set for the algorithm are mentioned in Table1. For PSO, we used $C_1 = C_2 = 1.496$ [35]. A, B are not required by PSO. The convergence of these four algorithms is shown in Figure 3. It is observed that single dimension change APSO (SDCAPSO) converges fast giving the minimum value of the objective function in a lesser number of iterations.

Table 1 Parameters of Proposed Algorithm

No.	Parameters	Value selected
1	Population size (N)	100
2	Nvars	13
3	MaxIter	650
4	A	0.3
5	B	0.5
6	W	0.9 at start 0.4 at end
7	Upper bound	0.5
8	Lower bound	-0.5

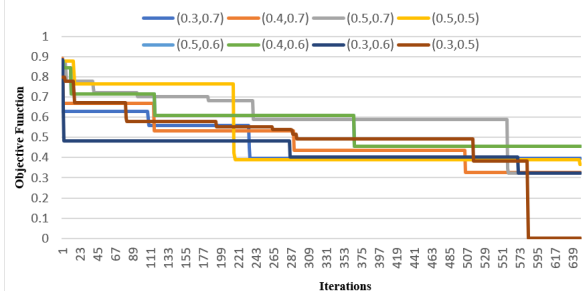**Figure 1** Effect of population size on convergence

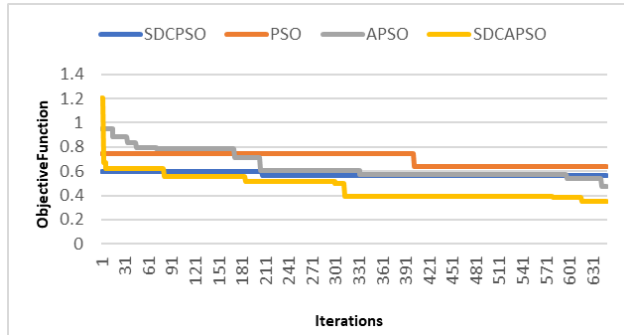**Figure 2** Effect of variation of (A, B) on the convergence

Figure 3 Convergence comparison

Stopping Criterion

The algorithm goes on searching for the optimized solution until the stopping condition is met. When the objective function value reaches the threshold value set in the algorithm, the algorithm stops running. The threshold used here is 0.4. So, the stopping condition is *ObjFunction < = 0.4*.

3.0 RESULTS AND DISCUSSION

To evaluate the performance of our proposed algorithm, we used the NOIZEUS database for noisy speech input. This database comprises thirty IEEE sentences produced by male and female speakers corrupted by real-world noises [37]. We used MATLAB R2019b for simulation on the system with Intel(R) Core i5- CPU at 1.20 GHz and 8 GB RAM. The performance of the proposed algorithm is compared with that of various other algorithms with respect to the objective speech quality measures as PESQ and segmental SNR. Perceptual Evaluation of Speech Quality (PESQ) is a test methodology employed for automated assessment of speech quality. It is the standard for objective voice quality testing. PESQ is the full-reference algorithm and analyzes the speech signal sample-by-sample after a temporal alignment of corresponding excerpts of reference and test signal. PESQ can be applied to provide an end-to-end (E2E) quality assessment for a network [38]. The PESQ score is computed as a linear combination of the average disturbance value and the average asymmetrical disturbance values and is given by the eqn. (14).

$$PESQ = b_0 + b_1 * D_{avg} + b_2 * A_{avg} \qquad (14)$$

$$where\ b_0 = 4.50, b_1 = -0.10, b_2 = -0.0309$$

Segmental SNR (segSNR) is the most known objective speech quality measure in the time domain. It calculates the average of the SNR values of short segments (frames). It is given by eqn. (15).

$$SNRseg = \frac{10}{M} \sum_{m=0}^{M-1} log_{10} \sum_{i=Nm}^{Nm+N-1} \frac{\sum_{i=1}^{N} x^2 (i)}{\sum_{i=1}^{N}(x(i) - y(i))^2} \qquad (15)$$

where *N* and *M* are the segment length and the number of segments respectively. Here, we used standard MATLAB code for the estimation of PESQ and segmental SNR.

Performance Comparison of PSO, APSO, SDCPSO and Proposed algorithm

To compare the performance of the above-mentioned four algorithms, we set the parameters of algorithms as specified before. The results are compared for three noise types namely babble, car, and airport with overall signal to noise ratio (SNR) of 5dB. Six sentences by male and five by female speakers are randomly selected as the input noisy sentences. A total of 33 (11*1*3) sentences are used to evaluate the objective measures. Perceptual evaluation of speech quality (PESQ) and segmental SNR are estimated in each case. For every noisy sentence, the algorithm is run 20 times to have the mean objective function value and the mean of the output parameter values. Figure 4a and Figure 4b give a percentage increase in PESQ and improvement in segmental SNR value for each algorithm respectively. For the proposed algorithm, we got a 17.45% improvement in PESQ for babble noise, 33.92% for car noise, and 14.96% for airport noise. Thus, for all the three types of noises improvement in PESQ is more for the proposed algorithm as compared with PSO, SDCPSO, and APSO. We got an increase in segmental SNR by 8.2dB for babble noise, 8.7dB for car noise, and 8.16dB for airport noise. Improvement in segmental SNR is more for car noise as compared to the rest of the three algorithms. For babble and airport noise, the performance of our algorithm is similar to APSO.

Performance Comparison with Other Algorithms

Six sentences by male and five by female speakers corrupted with babble, car, and exhibition noise having overall SNR of 0dB, 5dB, and 10dB are selected randomly from the NOIZEUS database as the input noisy sentences. Altogether 99(11*3*3) sentences are used for objective evaluation. The output PESQ result of the proposed algorithm is compared with the results of bnmf, MMSE, MMSE-PSO [33], and log-MMSE for all the three noise types at all SNR levels and the result of multi-level single-channel speech enhancement [34] for babble noise at 0dB and 5dB.
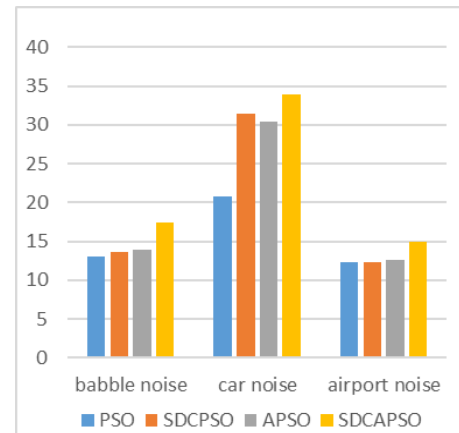

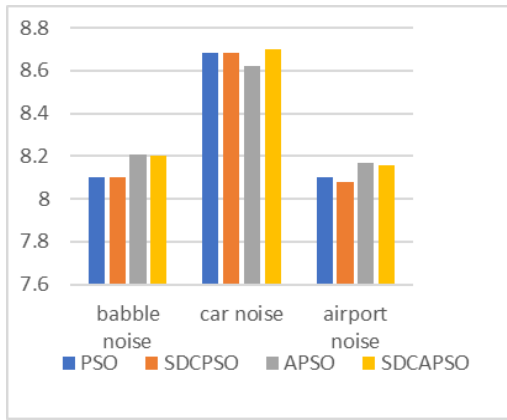**Figure 4a** Percent increase in PESQ

Figure 4b Improvement in SegSNR

Table 2 gives the output PESQ for these methods. It is observed that PESQ of the enhanced speech is more for the proposed algorithm compared with the other methods for all noise types and SNR levels. Figures 5, Figure 6 and Figure 7 give a comparison of output PESQ of these algorithms for babble, car, and exhibition noise. For 0dB babble noise, we got an average output PESQ of 1.9108, for 5dB babble it is 2.2812 and for 10dB it is 2.6234. In each of these cases, the PESQ of the output of the proposed algorithm is more compared to the other algorithms. Also, it is seen from figure 6 and figure 7 that the average output PESQ for the proposed algorithm is better compared to other algorithms for car and exhibition noise for 0dB and 5dB SNR cases. Table 3 gives the comparison between the output PESQ results of the proposed algorithm and the log-MMSE results. The percentage increase in PESQ is more in the case of the proposed algorithm for all SNR levels compared to log-MMSE. It is observed that the percentage increase in PESQ is more for babble and exhibition noise for 0dB and 5dB SNR levels as compared to car noise. Table 4 presents the comparison of improvement in segmental SNR of proposed algorithm and log-MMSE. Improvement in segmental SNR is more for 0dB and 5 dB but for 10dB, it is lesser than log-MMSE.

Table 2 Output PESQ for various algorithms

Noise type	Method	0 dB	5 dB	10 dB
Babble	Bnmf	1.70	2	2.25
	MMSE	1.72	2	2.3
	MMSE-PSO	1.75	1.895	2.45
	Multi-Level	1.849	2.1451	-
	Log MMSE	1.8630	2.2065	2.5973
	Proposed algorithm	**1.9108**	**2.2812**	**2.6234**
Car	bnmf	1.625	1.85	2.20
	MMSE	1.6	2.15	1.9
	MMSE-PSO	1.75	2.13	2.25
	Log MMSE	1.9221	2.3394	2.7090
	Proposed algorithm	**1.9375**	**2.3863**	**2.7322**
Exhibition	bnmf	1.3	1.9	2.2
	MMSE	1.31	1.7	2.1
	MMSE-PSO	1.75	1.95	2.4
	Log MMSE	1.6712	2.1703	2.5334
	Proposed algorithm	**1.7821**	**2.1857**	**2.5510**

Table 3 Comparison of Output PESQ of Proposed Algorithm with Log-MMSE

Noise	Percentage PESQ Increase					
	Log-MMSE			Proposed		
	0 dB	5dB	10dB	0dB	5dB	10dB
babble	11.66	13.50	12.91	**14.49**	**17.05**	**14.05**
car	21.68	31.48	26.03	**22.01**	**33.92**	**27.11**
exhibition	10.71	18.37	17.92	**13.75**	**39.13**	**18.74**

Table 4 Comparison of Seg SNR Results of Proposed Algorithm with Log-MMSE

Noise	SegSNR Increase					
	Log-MMSE			Proposed		
	0 dB	5dB	10dB	0dB	5dB	10dB
babble	6.57	6.28	4.24	**10.86**	**7.89**	2.43
car	10.08	8.54	6.66	**12.51**	**8.70**	3.42
exhibition	7.08	5.50	8.02	**11.43**	**6.15**	2.27

Figure 8 shows the convergence plot of the proposed algorithm for 5dB car noise. The algorithm converges to the threshold value in 20 iterations giving optimized values for the filter coefficients. Figure 9, Figure 10 and Figure 11 show the spectrograms of clean, noisy, and enhanced speech for different types of noises. Figure 8 shows the spectrogram of sentence spoken by male speaker with the babble noise at 0dB SNR. It is seen that the noise in the enhanced speech is reduced compared to the noisy speech which tells that the quality of enhanced speech is improved. Figure 10 gives the spectrogram of the sentence spoken by a female speaker with car noise. We get clear formants in the output with the reduction in noise. In Figure 11, a spectrogram of the sentence spoken by a male speaker with airport noise is shown. Figure12 gives a magnitude plot of the filter designed by the proposed algorithm. The multiple stopbands filter out the noise present in the input speech. It improves the segmental SNR value of the enhanced speech. With the smaller threshold values (0.001) kept in low-frequency bands in the formulation of the objective function, there is an improvement in PESQ values of the enhanced speech.
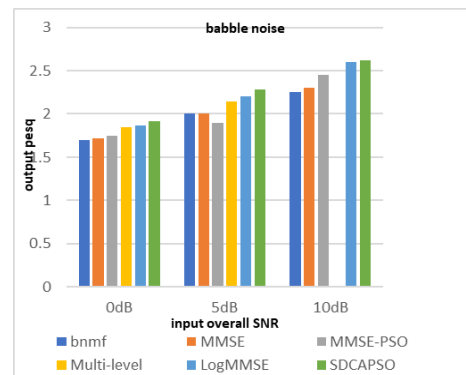

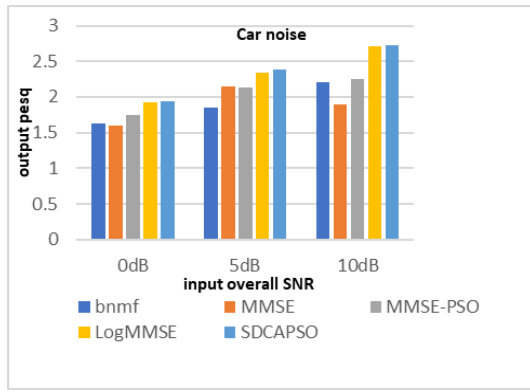**Figure 5** Comparison of output PESQ for babble noise

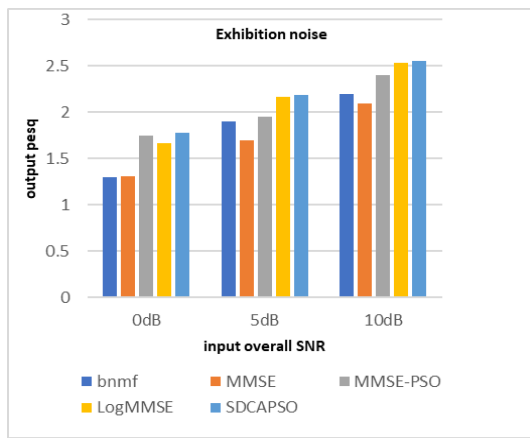**Figure 6** Comparison of output PESQ for car noise

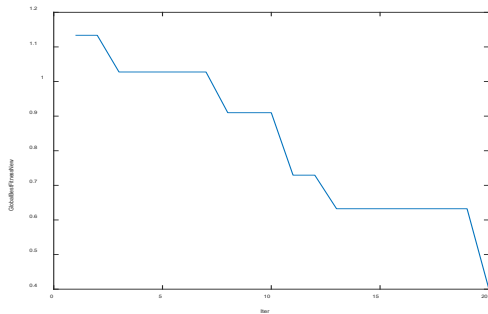**Figure 7** Comparison of output PESQ for exhibition noise

Figure 8 Convergence of the proposed algorithm for Sp12 with 5dB car noise
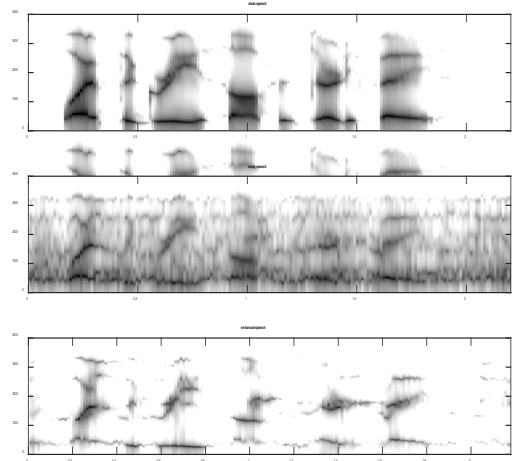

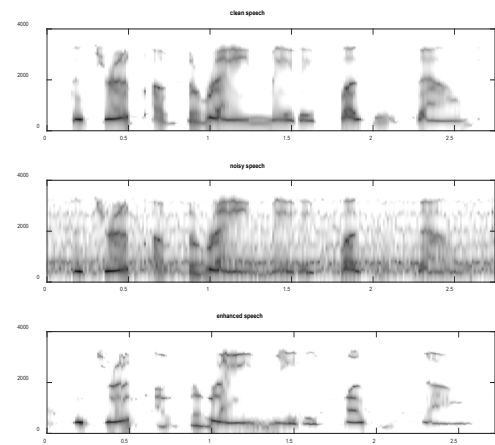**Figure 9** Spectrogram of Sp05 (male speaker) with 0dB babble noise PESQ = 2.1248

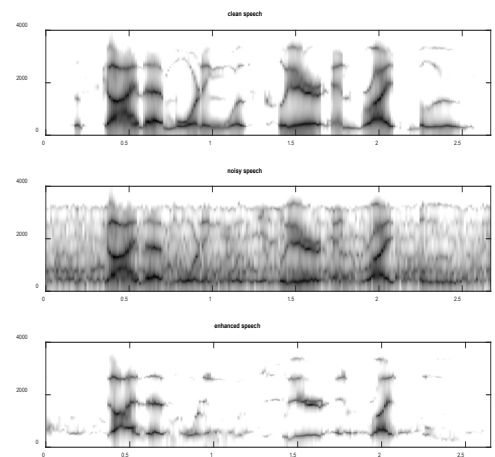**Figure 10** Spectrogram of Sp12 (female speaker) with 5db car noise PESQ = 2.3794

Figure 11 Spectrogram of Sp10 (male speaker) with 5dB airport noise PESQ = 2.4668
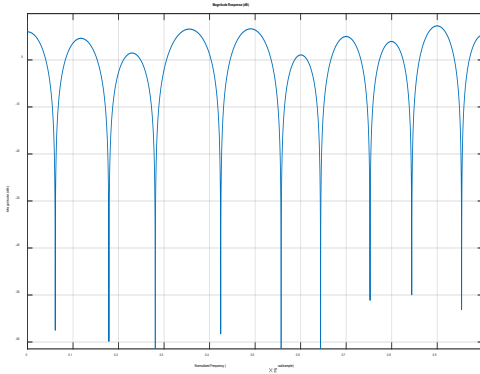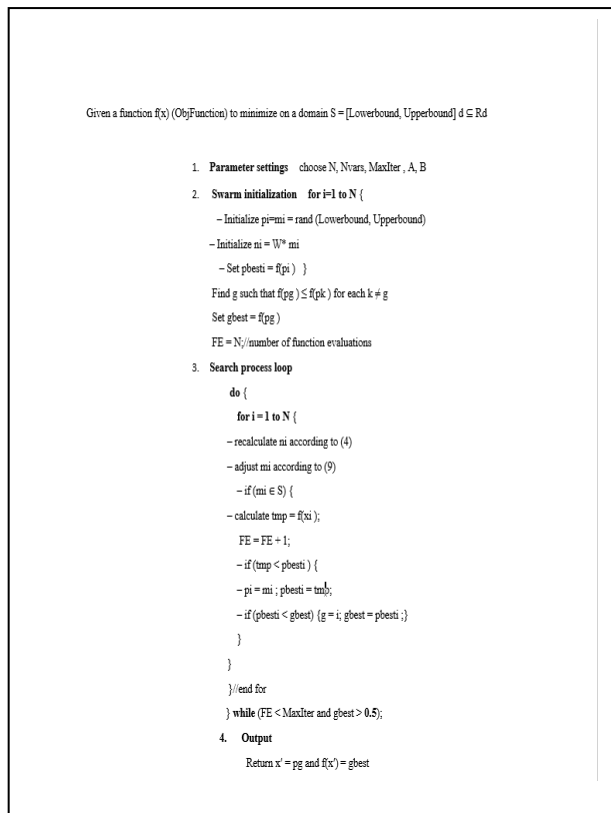
Figure 12 Magnitude plot of the filter

Scheme 1 gives the pseudocode of the proposed algorithm.

Scheme 1 Pseudo code of SDCAPSO

4.0 CONCLUSION

In this work, we implemented a single-channel speech enhancement system. In the proposed algorithm we used log-MMSE for pre-processing the noisy speech and then a single dimension change APSO is implemented to filter the signal. This technique gives an improvement in PESQ indicating improvement in the intelligibility of input noisy speech. For 5dB

babble noise, we got 17.05% increase in PESQ, for 5dB car noise it is 33.92 %, for 5dB airport noise it is 14.96 % and for 5dB exhibition it is 39.13 %. Segmental SNR increases and the spectrograms show that there is improvement in the quality of the noisy speech. The convergence rate of the proposed algorithm is higher compared to PSO and APSO giving optimization in a lesser number of iterations.

Evolutionary algorithms like differential evolution, cuckoo search algorithm, bat algorithm, and hybrid algorithms may be used for filtering of noise in speech for single-channel speech enhancement in the future for improving intelligibility with minimum speech distortion.

References

[1]   Kondaz, A. M., 2004. *Digital speech coding for low bit rate communication systems,* Second Edition, (John Wiley and Sons) DOI: https://doi.org/10.1002/0470870109
[2]   Loizou, P. C. 2013. *Speech Enhancement: Theory and Practice*, Second Edition CRC Press DOI: https://doi.org/10.1201/b14529
[3]   Hu, Yi and Loizou, P.C. 2006. Subjective Comparison of Speech Enhancement Algorithms. Department of Electrical Engineering, University of Texas at Dallas Richardson, Texas. 1-4244-0469-X/06 IEEE DOI: 10.1109/icassp.2006.1659980
[4]   Boll, S. 1979. Suppression of acoustic noise in speech using spectral subtraction, *IEEE Transactions on acoustics, speech, and signal processing*, 27(2): 113–120. DOI: https://doi.org/10.1109/TASSP.1979.1163209
[5]   Berouti,M. , Schwartz,R. , and Makhoul,J.1979.Ehancementof speech corrupted by acoutic noise,*IEEE International Conference on Acoustics, Speech, and Signal Prcessing, ICASSP '79*, 4: 208-211 DOI: 10.1109/ICASSP.1979.1170788
[6]   Zadeh, L. 1950. Frequency analysis of variable networks, *Institute of Radio Engineering.* 38: 291-299. DOI: https://doi.org/10.1109/JRPROC.1950.231083
[7]   Atlas,L. 2003.Modulation spectral transforms: Application to speech separation and modification, University of Washington, Washington, WA
[8]   Paliwal, K. Wojcicki, and Schwerin, B. 2010. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain, *Speech Communication,* 52(5):450–475. DOI: https://doi.org/10.1016/j.specom.2010.02.004
[9]   Zang, Yi 2012. Modulation domain processing and speech phase spectrum in speech enhancement, A Dissertation Presented to the Faculty of the Graduate School at the University of Missouri-Columbia
[10]   Wang, Y. 2015. Speech enhancement in the modulation domain, PhD thesis, Imperial College London
[11]   Dionelis N. and Brookes M. 2017. Modulation domain speech enhancement using Kalman Filter with a Bayesian update of speech and noise in the log spectral domain, 978-1-5090-5925-6/*IEEE Proceeding Hands-Free Speech Communication and Microphone Arrays, HSCMA.* 111 - 115 DOI: https://doi.org/10.1109/HSCMA.2017.7895572
[12]   Wang, Y., and Brookes M. 2018. Model-Based Speech Enhancement in the Modulation Domain, *IEEE/ACM Transaction on Audio, Speech and Language Processing,* 26(3): 580–594. DOI: https://doi.org/10.1109/TASLP.2017.2786863
[13]   Widrow, B. and Stearns, S.D. 1985. Adaptive Signal Processing, Prentice-hall Englewood Cliffs, NJ.
[14]   Mohammed, J.R. 2007. A new simple adaptive noise cancellation scheme based on ale and NLMS filter, *Proceedings of the 5th Annual Conference on Communication Networks and Services Research*, May 14-17, IEEE Xplore Press, Frederlcton, NB, Canada, 245-254. DOI: https://doi.org/10.1109/CNSR.2007.4
[15]   Gorriz, J.M., Ramırez, J., Cruces-Alvarez, S., Puntonet, C.G. and Lang, E.W. et al.: "A novel LMS algorithm applied to adaptive noise cancellation, *IEEE Signal Process. Lett.,* 16: 34-37.DOI: https://doi.org/10.1109/LSP.2008.2008584

[16] Shynk, J. J. 1989. Adaptive IIR Filtering, *IEEE ASSP Magazine*, 4–21 DOI: https://doi.org/10.1109/53.29644

[17] Krusicnski, D.J. and Jenkins, W.K. 2003. Adaptive Filtering Via Particle Swarm Optimization, *Proc. 37'Asilomar Conf on Signals, Systems, and Computers*. DOI: https://doi.org/10.1109/acssc.2003.1291975

[18] Krusienski, D. J. and Jenkins, W. K. 2004. Particle Swarm Optimizationfor Adaptive IIR Filter Structures, 0-7803-8515-2/04/2004, *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753).* DOI: https://doi.org/10.1109/cec.2004.1330966

[19] Yang X. S. 2008. Nature-Inspired Metaheuristic Algorithms, Luniver Press.

[20] Yang, Xin-She, Deb, S. and Fong, S. 2011. Accelerated Particle Swarm Optimization and Support Vector Machine for Business Optimization and Applications, *Networked Digital Technologies (NDT2011), Communications in Computer and Information Science,* Vol. 136, Springer, 53–66.DOI: https://doi.org/10.1007/978-3-642-22185-9 6

[21] Mandal, S., Ghoshal, S., Kar, R., Mandal, D. 2012. Design of optimal linear phase FIR high pass filter using craziness-based particle swarm optimization technique, *Journal of King Saud University – Computer and Information Sciences*, 24, 83–92. DOI: https://doi.org/10.1016/j.jksuci.2011.10.007

[22] Mandal, S., Ghoshal, S., Kar R., Manda,l D. 2012.Craziness based Particle Swarm Optimization algorithm for FIR band stop filter design, *Swarm and Evolutionary Computation*, 7: 58–64. DOI: https://doi.org/10.1016/j.swevo.2012.05.002

[23] Aggarwal, A., Rawat, T., Upadhyay,D. 2016. Design of optimal digital FIR filters using evolutionary and swarm optimization techniques, *International Journal of Electronics and Communication* (AEÜ), 70: 373–385.DOI: https://doi.org/10.1016/j.aeue.2015.12.012

[24] Lim, W. H. and Nor A. M. I. 2015. Particle Swarm Optimization with Improved Learning Strategy, *Journal of Engineering Science*, 11: 27–48.

[25] Zhao F. 2016. Optimized Algorithm for Particle Swarm Optimization, *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering,* 10(3): 91-95. DOI: https://doi.org/10.1155/2016/3968324

[26] Xu, G., Cui,Q., Shi,X., Ge, H., Zhan,Z., Lee, H. P., Liang,Y., Tai,R., Wu,C. 2019. Particle swarm optimization based on dimensional learning strategy, *Swarm and Evolutionary Computation,* 45: 33–51. DOI: https://doi.org/10.1016/j.swevo.2018.12.009

[27] Fajr, R., and Bouroumi, A. 2017. An Improved Particle Swarm Optimization Algorithm for Global Multidimensional Optimization, *Journal of Intelligent Systems*, 29(1): 127–142. DOI: https://doi.org/10.1515/jisys-2017-0104

[28] Zhang, Y., Wang,S., and Ji,G. 2015. Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications, *Mathematical Problems in Engineering*, Article ID 931256, 38 pages. DOI: https://doi.org/10.1155/2015/931256

[29] Prajna, K., Rao, G.S.B., Reddy, K. V. V. S. 2014. A New Dual Channel Speech Enhancement Approach Based on Accelerated Particle Swarm Optimization (APSO), *International Journal of Intelligent Systems and Applications.* DOI: https://doi.org/10.5815/ijisa.2014.04.01

[30] Prajna, K., Rao, G.S.B., Reddy, K. V. V. S., Maheswari, R. U., 2015 .A new approach to dual channel speech enhancement based on hybrid PSOGSA, *International Journal of Speech Technology*, 18: 45–56. DOI: https://doi.org/10.1007/s10772-014-9245-5

[31] Geravanchizadeh, M.,Osgouei S. G., 2015. A New Shuffled Sub-Swarm Particle Swarm Optimization Algorithm for Speech Enhancement, *Journal of Advances in Computer Engineering and Technology*, 1(1): 43-50

[32] Sandeep Kumar, 2020.Directed Searching Optimization-Based Speech Enhancement Technique, *Fluctuation and Noise Letters,* 2050035, World Scientific Publishing Company. DOI: https://doi.org/10.1142/S0219477520500352

[33] Selvi,R. S.,Suresh G.R. 2015: Hybridization of spectral filtering with particle swarm optimization for speech signal enhancement, *International Journal of Speech Technology*, 19(1): 19-31 DOI: https://doi.org/10.1007/S10772-015-9317-1

[34] Lavanya T., Nagarajan T., and Vijayalakshmi P. 2020. Multi-level Single-Channel Speech Enhancement Using a Unified Framework for Estimating Magnitude and Phase spectra, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 28: 1315-1327. DOI: https://doi.org/10.1109/TASLP.2020.2986877

[35] Kennedy J., and Eberhart R. 1995. Particle swarm optimization, *Proceedings of the IEEE International Conference on Neural Networks*, 4: 1942–1948.

[36] Wei X., Anyu Li., Boya S., and Zhao J. 2018. A Novel Design of Sparse FIR Multiple Notch Filters with Tunable Notch Frequencies, *Mathematical Problems in Engineering*, 2018, Article ID 3490830. DOI: https://doi.org/10.1155/2018/3490830

[37] Hu, Y. and Loizou, P. 2007. Subjective evaluation and comparison of speech enhancement algorithms, *Speech Communication*, 49: 588–601. DOI: https://doi.org/10.1016/j.specom.2006.12.006

[38] Rix A.W., Beerends G. J., Holliar M.P. 2001. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs, *IEEE International conference on Acoustic, Speech and Signal Processing proceedings* (Cat. No.01CH37221). DOI: https://doi.org/10.1109/ICASSP2001.941023.