International Journal of Scientific and Technical Advancements
ISSN: 2454-1532

Genetic Algorithm based Speech Enhancement in Noisy Environment

Neha Jain[1#], Sandeep Singh Gill[1], Parveen Kumar Lehana[2]
[1]ECE Department, GNDEC, Ludhiana, Punjab, India
[2]Department of Electronics, University of Jammu, Jammu, India
[#]Email address: runjhun.neha@gmail.com

Abstract—Speech is a dynamic biomedical signal used for the ease of communication. Effective speech communication requires the preservation of the important features such as quality, intelligibility, and pleasantness. However, the presence of noise in the background usually degrades one or more of these features resulting into poor perception of the speech. This paper presents the development and the performance analysis of a GA based speech enhancement technique for reducing spectral distortion in the speech degraded by white Gaussian noise present in the background. The proposed algorithm utilizes the robust platform of Harmonic plus Noise model (HNM) for the analysis, modification, and resynthesis of the degraded speech. The concepts of genetics have been used for optimizing the spectral parameters of the degraded speech in order to minimize the Log Spectral Distance (LSD) between the original recorded and the processed speech. LSD was selected as the fitness function for GA based processing. GA was initialized with the population of 10 chromosomes, each consisting of 16 genes representing the parameters required for spectral modification of the input degraded speech. The speech was resynthesized using modified spectral parameters. About 200 generations were estimated to be sufficient for estimating the optimum values of the spectral parameters necessary for speech enhancement. The results show the average reduction in the output LSD at 0 dB, 5 dB, and 10 dB relative to the input LSD as 52.03%, 47.96%, and 37.99%, respectively. The amount of reduction in LSD confirms the effectiveness of the proposed algorithm for the enhancement of the speech corrupted by the white Gaussian noise.

Keywords— Genetic algorithm; Harmonic plus noise model; Log spectral distance; white Gaussian noise

I. INTRODUCTION

Speech is a time varying acoustic signal. It carries information about the speaker's identity, gender, emotional state along with the linguistic information. It serves as one of the most natural and convenient way of exchanging information among human beings [1]. The sub glottal system containing lungs, trachea, and bronchi initiates the process of speech production by the expulsion of a constant velocity air stream from the lungs. The vocal cords convert the constant velocity flow into excitation signal needed for initializing the vocal tract acting as a filter. The excitation is then modulated by the articulators in the vocal tract to produce different types of sounds. The voiced sounds are produced due to vibration of vocal cords. In case of voiced sounds, the excitation signal is produced when the air is forced to flow through the glottis with appropriate adjustment in tension of the vocal cords. The unvoiced sounds, containing non periodic components of speech, are produced without vibrating the vocal cords, but the excitation is produced when the air flow undergoes constriction at some point in the vocal tract [2]. The output speech is radiated at the lips and propagates to the listeners' ear through the air. The speech signals produced by the human beings many a times need to be processed to improve its perception for the listeners particularly in the presence of background noise. This requires the understanding of hearing physiology.

The three main parts of the ear are outer, middle, and inner ear (Fig. 1). The outer ear consists of the pinna, the ear canal, and the ear drum. The sound from the air is collected by the pinna and transmitted to the ear drum through the ear canal. The sound reaching the ear drum sets it into the vibrations. These vibrations are then transmitted to the ossicles in the middle ear. It consists of three hearing bones- the stapes, the incus, and the malleus. The inner ear known as cochlea is positioned behind the oval window and contains fluid inside it. The primary function of the ossicles is to match the low impedance of air in the outer ear to the high impedance of the fluid inside the cochlea. The cochlea also contains the inner hair cells and the outer hair cells. These hair cells act as sensory cells. The vibration of the oval window results into the movement of hair cells. The outer hair cells dampen the loud sound and amplify the soft sound. The inner hair cells transfer the sound to the auditory nerves for perception in the brain [3].

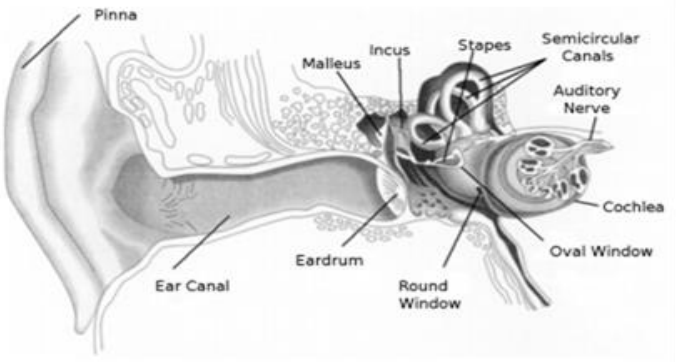

Fig. 1 The outer, the middle, and the inner ear [3]

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

Improper functioning of the ear leads to hearing loss. It is a common disorder affecting 10% - 15% of the world's population. The hearing loss reduces the sensitivity at various frequencies which in turn reduces the perception of the speech. The hearing loss is of two types that is sensorineural and conductive. The main reason for the conductive hearing loss is the obstruction in the transmission of sound inside the ear caused either by the abnormal growth of bones or the accumulation of wax. The sensorineural hearing loss on the other hand is mainly due to the damage of the inner and the outer hair cells in the cochlea because of exposure to excessive noise, aging, and ear disease. The sensorineural loss or noise induced hearing impairment usually results in loss at high frequency. This loss is the main reason for hearing impairment in 50% of the population above the age of 70 [4]. Further, hearing loss may be unilateral or bilateral. People with unilateral hearing loss require considerably more effort for understanding the speech in noisy environment and localization of sound [5].

The presence of noise degrades one or more important features of the speech – quality, intelligibility, and listening comfort necessary for its perception [6]. The struggle of understanding speech in noise is more severe for the listeners with hearing loss due to their elevated hearing threshold which is reported to be 12 dB more than the normal hearing listeners resulting into limited audibility [7], [8]. The perception of speech in noisy background for people with hearing loss could be enhanced with the use of two classical approaches namely directional microphones and noise reduction algorithms. The directional microphones are benefitted from the spatial difference between speech and noise to enhance the speech coming from a specific direction. On the other hand, spectral and temporal difference between the noise and the speech is utilized by the noise reduction algorithms for noise suppression. The main limitation of the directional microphones is their reduced directivity in reverberant environments, higher internal noise, and reduced gain in low frequency regions. The size and the number of microphones restrict their use in deep ear hearing aids. The main limitation of the noise reduction algorithm is their inability to differentiate between the desired speech signal and the unwanted noise in case of spectral overlap [9], [10].

This paper presents the development of a GA based algorithm for reducing spectral distortion in the speech degraded by the white Gaussian noise present in the background. The proposed algorithm processes the input degraded speech using an analyses-modification-resynthesis system. The concepts of genetics have been utilized for the dynamic optimization of spectral parameters of the degraded speech such that the LSD between the original recorded and the processed speech is minimized. The GA was selected for optimization because of its ability to find global optima with less prior knowledge of the search space. The modified spectral parameters have been used for the re-synthesis of the speech. Analysis/synthesis framework of Harmonic plus Noise Model (HNM) has been utilized for the analysis of the degraded speech and re-synthesis of the modified speech. The details of HNM are provided in Section II. The methodology of the proposed work is explained in Section III. The results and conclusion is presented in Section IV and Section V, respectively.

II. HARMONIC PLUS NOISE MODEL

Harmonic plus noise model is a robust analysis/synthesis model that divides the speech spectrum into Harmonic and noise part [11], [12], [13], [14], [15].

$$s(t) = s'(t) + n'(t) \tag{1}$$

The quasi periodic components of speech are represented by the harmonic part while the non-periodic components of speech such as stops, fricatives, and aspirates are simulated by the noise. The two bands are separated by the maximum voiced frequency F_m. Speech signals below F_m are harmonically related sine waves represented in as

$$s'(t) = \sum_{k=-L(t)}^{L(t)} A_k(t) e^{jk\omega_0(t)} \tag{2}$$

Here, $L(t)$ represents the number of harmonics included in the harmonic part of speech, $\omega_0(t)$ is the fundamental frequency, and $A_k(t)$ represents complex amplitudes as function of time. The spectrum above F_m appears to be noisy and is modelled in as white Gaussian noise $b(t)$ filtered by an all pole normalized filter whose amplitude is modulated by an energy envelope function given as

$$n'(t) = w(t)[h(t,\tau) * b(t)] \tag{3}$$

In HNM analysis, speech signal is divided into frames and analysis is carried out frame by frame. The decision of voiced and unvoiced frames is taken by the voice activity detector and the maximum voiced frequency F_m is calculated. The pitch synchronous analysis and synthesis in HNM requires the analysis window of two pitch periods. The pitch period in case of voiced frame, refers to the local pitch period while it is assumed as 10 ms to account for the non-periodic nature of unvoiced frame. The important parameters of voiced part like amplitudes and phases of all pitch harmonics up to F_m are calculated by analysing voiced frames at each GCI under the assumption that analysis window extends from previous GCI to next GCI. In case of voiced frame, the harmonic part of speech synthesized using (2) is subtracted from the original speech signal to obtain noise part. The entire unvoiced frame is considered as noise part. Noise part of the voiced and the unvoiced frame is subjected to LPC analysis to obtain LPC coefficients and energy envelope which is important for the perception of stops, fricatives, and aspirates. The length of analysis window is kept two local pitch periods (same as analysis of harmonic part) for the noise part of the voiced frame and 20 ms with 10 ms overlap for the unvoiced frame. The parameters extracted from the HNM analysis are

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

interpolated at frame boundaries and may be modified before synthesis. Synthesis of harmonic part is carried out using (2). Noise part is synthesized using (3) where the all pole filter is defined by LPC coefficients and energy envelope estimated during HNM analysis of the speech. Finally, the synthesised speech signal in (4) is given as the addition of synthesized harmonic and synthesized noise part.

$$\hat{s}(t) = s'(t) + n'(t) \quad (4)$$

One of the factors contributing towards the robustness of HNM model is its ability to synthesize syllables and passages spoken with different articulatory position and styles with high quality and naturalness [13], [14]. Speech enhancement can be efficiently carried out in low signal-to-noise ratio environment by using analysis/synthesis framework of HNM. The HNM assumes the speech to be composed of voiced part that contains only harmonically related sounds and unvoiced part containing only random signals. The effective and flexible decomposition of speech along with its compact parameter requirement increases its suitability for speech enhancement [16]. Further, speech having harmonic characteristics, the real world low frequency components of noise can be eliminated from voiced speech, eliminating the remnant noise. HNM is reported as a robust model for analysis, modification, and synthesis of speech because of its ability to separately control and modify the noise and voice bands in the frame-wise spectrum of speech [17].

III. METHODOLOGY

In the present work, an algorithm is developed to enhance the speech degraded by the additive spectrally shaped white Gaussian noise by minimising the spectral distortion. The constituents blocks of the algorithm and flow of information among them is shown in Fig. 2. The speech material used for the investigation comprises of total 10 sentences taken from TIMIT database and self-recorded speech in male and female voices given by subject codes (S1V1-S10V10). The text material used for the recording consists of phonetically balanced sentences in Hindi taken from TIFR (Tata Institute of Fundamental Research) database. The recording was done in an acoustically treated room using a high quality Sony GX-400 voice recorder with sampling frequency of 16 KHz at 16 bit quantization. White Gaussian noise was added at different values of SNR to the original recorded speech to account for the quantization noise of the digital devices like hearing aids, amplifiers, electronic components, etc., present even in the quietest environment. The unprocessed speech was divided into frames and each frame was classified as either voiced or unvoiced. For voiced frames, important parameters of speech like harmonic magnitudes, harmonic phases, pitch frequency, and maximum voiced frequency were extracted. The extracted parameters were optimized using genetic algorithm to minimize the LSD between the processed and the recorded speech for the efficient spectral modification of the unprocessed speech. GA was initialized with a population of 10 chromosomes each made up of 16 genes representing the parameters required for the spectral modification of the unprocessed speech and hence its enhancement.

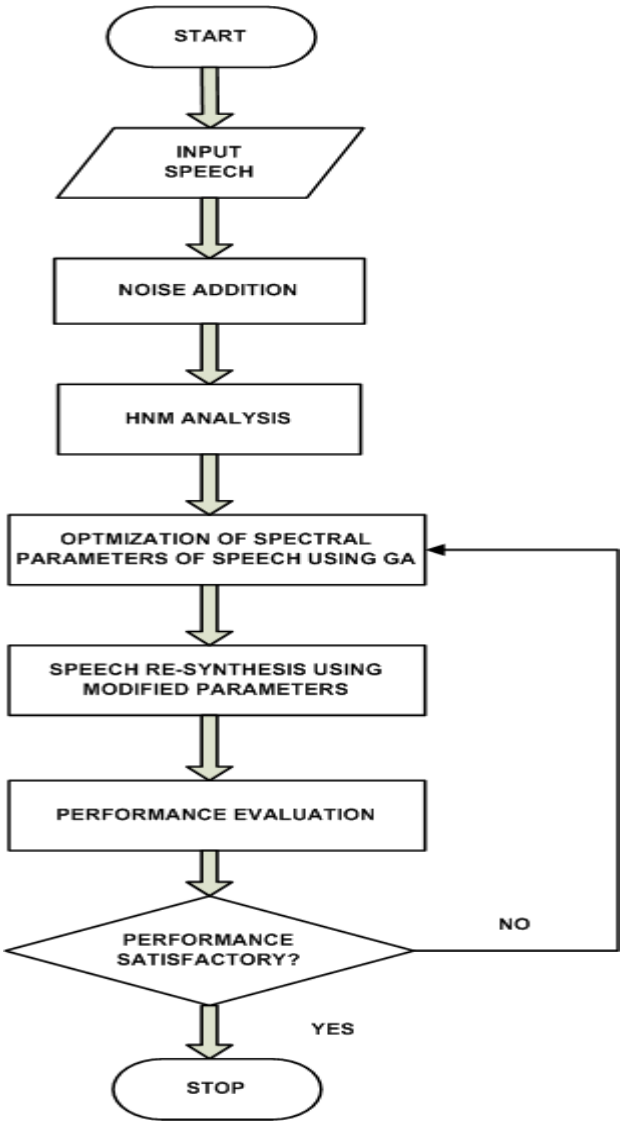

Fig.2 Proposed algorithm

LSD was selected as the fitness function that needs to be minimized for reducing spectral distortion. The mathematical formulation of LSD reported in [16] has been used

$$LSD = \frac{1}{j}\sum_{l=0}^{j-1}\left\{\frac{1}{N/2+1}\sum_{k=0}^{N/2}\left[10\log_{10} X(k,l) - 10\log_{10} \hat{X}(k,l)\right]^2\right\}^{1/2} \quad (5)$$

where, N is the frame length, j represents the total number of frames and l is the frame number. $X(k,l)$ and $\hat{X}(k,l)$ are the short time Fourier transformation of the recorded speech and the processed speech, respectively. In each generation, the genes of each chromosome were used to process the spectrum

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

ISSN: 2454-1532

of unprocessed speech. LSD was calculated between the spectrum of the recorded and the processed speech obtained for each chromosome using (5). The chromosomes resulting into least LSD were selected to form a new generation by applying nature inspired genetic operators like selection, crossover, replacement, and mutation. Iterations were carried out till there was no further improvement in LSD which was considered as the stopping criteria. The best chromosome of the last generation resulted into the optimum value of genes for processing the spectral parameters obtained in HNM analysis of the unprocessed speech to reduce the spectral distortion. The modified spectral parameters were linearly interpolated to determine the value at each sample within the frame. Noise part was resynthesized using LPC platform. Finally, the enhanced speech signal was obtained by adding harmonic and noise parts. Towards this end investigations were carried out to estimate the number of generations required for minimising spectral distortion and to analyse the effect of white Gaussian noise on the estimation of optimal spectral parameters. Reduction in LSD was considered as performance measure for the evaluation of the proposed algorithm.

IV. RESULTS

The input (unprocessed) speech was processed using the algorithm presented in Section III. The LSD between the recorded and the unprocessed speech spectrum was taken as the input LSD. The LSD between the recorded and the processed speech spectrum was taken as the output LSD. This section investigates the number of generations needed in the GA based processing to minimize LSD and the percentage reduction in the output LSD in the presence of white Gaussian noise in the background at different values of SNR. The variation of output LSD with respect to the number of generations of GA at three different SNR values of 0dB, 5dB, and 10dB for four different speakers with subject codes (S1V1-S4V4) is shown in Fig. 3 to Fig. 6.

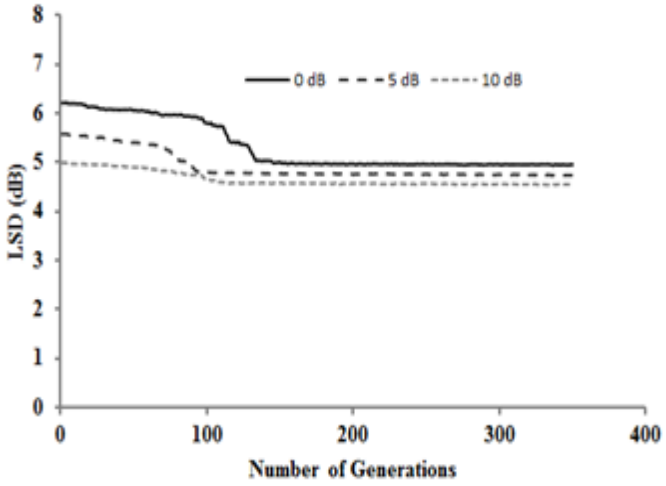

Fig. 3 Variation of LSD with respect to generations for subject code S1V1

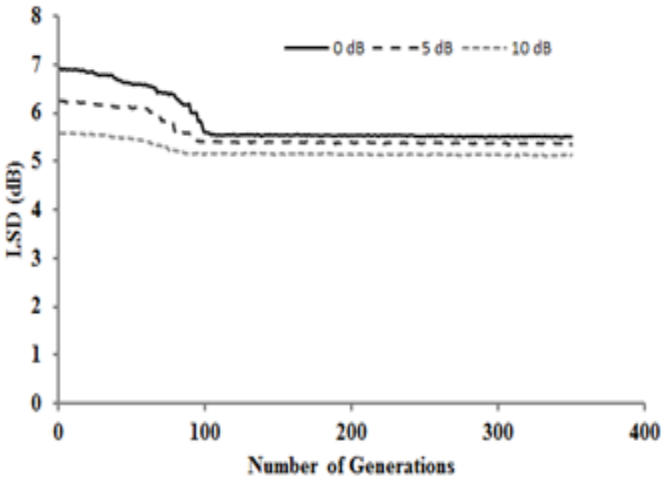

Fig. 4 Variation of LSD with respect to generations for subject code S2V2

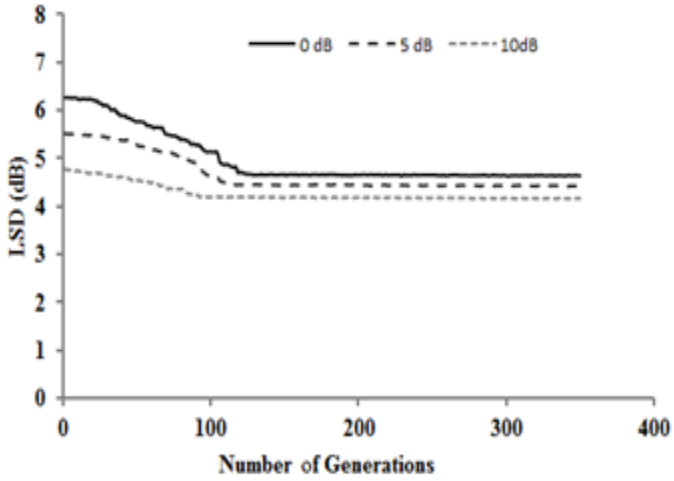

Fig. 5 Variation of LSD with respect to generations for subject code S3V3

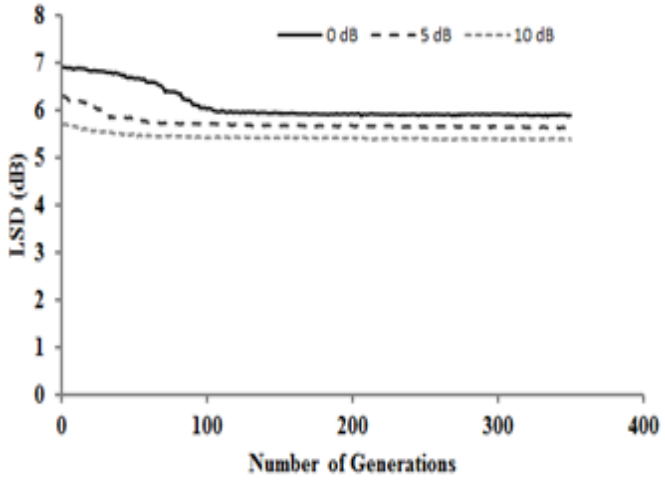

Fig. 6 Variation of LSD with respect to generations for subject code S4V4

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

It may be observed that after almost 200 generations, there is no considerable improvement in the LSD for all the three SNR values of 0 dB, 5 dB, and 10 dB. 200 generations were estimated to be sufficient for the convergence of GA. Also, output LSD decreases with the increase in the value of SNR. Similar results were observed for the remaining six subjects with subject code (S5V5-S10V10). The percentage reduction in the output LSD relative to the input LSD at different SNR values was used as a performance measure to investigate the effect of white Gaussian noise in the background on the estimation of spectral parameters. The input and output LSD for ten different sentences of ten different speakers at 0 dB, 5 dB, and 10 dB SNR values along with the respective percentage reduction in the output LSD is shown in Table 1. The comparison between mean input LSD and mean output LSD at 0 dB, 5 dB, and 10 dB SNR values along with the standard deviation is shown in the Table 2 and plotted as histograms in Fig. 7. The average reduction in the output LSD at 0 dB, 5 dB, and 10 dB relative to input LSD has been observed as 52.03%, 47.96%, and 37.99%, respectively. Analysis showed that both the output LSD and the input LSD decrease with increase in SNR. However, the output LSD is lesser than that of the input LSD at all the three SNR values.

Table 5.1 Comparison of input and output LSD at different SNR values for subject code (S1V1-S10V10)

Subject Code	Input SNR (dB)	LSD (dB)		Percentage reduction in output LSD (%)
		Input	Output	
S1V1	0	11.3640	4.6269	**59.28**
	5	10.0417	4.4214	**55.97**
	10	7.3495	4.2644	**41.98**
S2V2	0	11.1681	4.9516	**55.66**
	5	9.8354	4.7432	**51.77**
	10	8.2115	4.6396	**43.50**
S3V3	0	11.0408	4.9946	**54.76**
	5	9.6519	4.6807	**51.50**
	10	7.9383	4.6063	**41.97**
S4V4	0	12.2054	5.5111	**54.85**
	5	10.7980	5.3700	**50.27**
	10	8.0050	5.2320	**34.64**
S5V5	0	11.0107	6.1098	**44.51**
	5	9.7458	5.6119	**42.41**
	10	8.0986	5.4297	**32.96**
S6V6	0	11.5624	5.8983	**48.99**
	5	10.2689	5.6455	**45.02**
	10	8.6054	5.3920	**37.34**
S7V7	0	11.0784	5.3211	**51.96**
	5	9.8239	5.0786	**48.31**
	10	8.2731	4.8999	**40.77**
S8V8	0	11.4563	6.0440	**47.26**
	5	10.0877	5.9181	**41.33**
	10	8.4913	5.7834	**31.89**
S9V9	0	11.5414	5.6465	**51.08**
	5	10.1544	5.4904	**45.94**
	10	8.5092	5.2989	**37.73**
S10V10	0	11.4355	5.5181	**51.74**
	5	10.1323	5.3630	**47.07**
	10	8.4630	5.2610	**37.83**

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

Table 2 Comparison between mean input and mean output LSD at different SNR values

SNR	Mean LSD		Standard Deviation S_{DEV}		Reduction in output LSD (%)
	Input	Output	Input	Output	
0 dB	11.386	5.462	0.355	0.493	**52.03%**
5 dB	10.054	5.232	0.329	0.485	**47.96%**
10 dB	8.195	5.081	0.374	0.464	**37.99%**

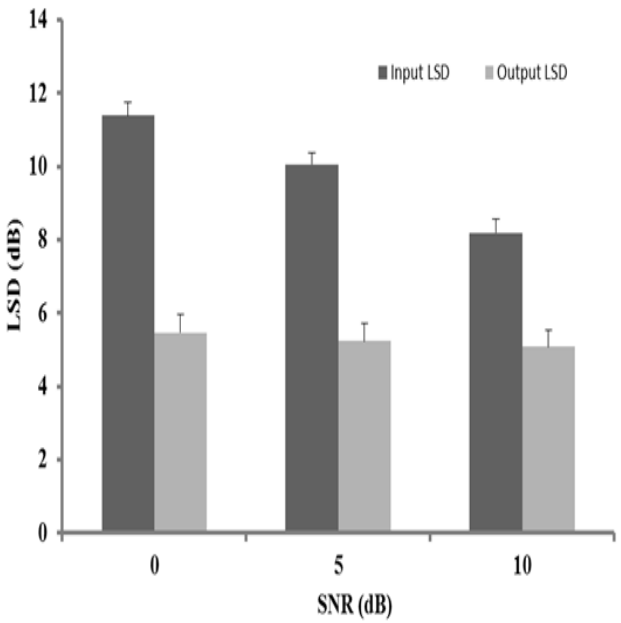

Fig. 7 Histogram representation of the mean input and output LSD with respect to SNR

I. CONCLUSION

Speech is an important means of the communication among human beings. However, the presence of noise in the background affects perception of speech by normal hearing listeners and more severely for the listeners with hearing impairedness. In this paper, a GA based algorithm has been developed and investigated for reducing spectral distortion introduced in the speech by the additive white Gaussian noise present in the background. For investigations, the speech material consisted of total ten sentences from TIMIT and TIFR databases. Investigations showed that about 200 generations are sufficient to minimize the LSD between the processed and the original recorded speech for satisfactory quality of the output. The average reduction in the output LSD at 0 dB, 5 dB, and 10 dB relative to the input LSD has been observed as 52.03%, 47.96%, and 37.99%, respectively, indicating the capability of the proposed algorithm in enhancing the noisy speech. The outcome of the research may be beneficial for the development of smart hearing aids.

REFERENCES

[1] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, and R. Rose, "Automatic speech recognition and speech variability: A review," *Speech Communication*, vol. 49, no. 10-11, pp. 763–786, 2007.

[2] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1978.

[3] L. C. Araújo, T. N. Magalhaes, D. P. Souza, H. C. Yehia, & M. A. Loureiro,"A brief history of auditory models," *10 Simpósio Brasileiro de Computaç ao Musical*, no. 1, 2005.

[4] J. L. Russell, H. S. Pine, D. L. Young, "Pediatric cochlear implantation: expanding applications and outcomes," *Pediatric Clinics of North America*, vol. 60, no. 4, pp. 841–863, 2013.

[5] J. E. Lieu, "Speech language and educational cons-equences of unilateral hearing loss in children," *Arch Otolaryngol Head Neck Surg*," vol. 130, no. 5, pp. 524-530, 2004.

[6] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: A regenerative approach," *Speech Communication*, vol. 10, no. 1, pp. 45-47, 1991.

[7] D. O' Shaughnessy, "Enhancing speech degraded by additive background noise or interfering speakers," *IEEE Communication Magazine*, vol. 27, no. 2, pp. 46-52, 1989.

[8] K. H. Arehart, J. H. Hansen, S. Gallant, and L. Kalstein, "Evaluation of an auditory masked threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners," *Speech Communication*, vol. 40, no. 4, pp. 575–592, 2003.

[9] K. Chung, "Challenges and Recent Developments in Hearing Aids: Part I. Speech Understanding in Noise," Microphone Technologies and Noise Reduction Algorithms," *Trends in Amplification*, vol. 8, no. 3, pp. 83–124, 2004.

[10] T. Jiang, R. Liang, Q. Wang, and Y. Bao, "Speech Noise Reduction Algorithm in Digital Hearing Aids Based on an Improved Sub-band SNR Estimation", Circuits, Systems, and Signal Processing, vol. 37, no. 3, pp. 1243–1267, 2017

[11] Y. Stylianou, "Applying the Harmonic Plus Noise Model in Concatenative Speech Synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 1, pp. 21-29, 2001.

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

ISSN: 2454-1532

[12] Y. Pantazis and Y. Stylianou, "Improving the modeling of the noise part in the harmonic plus noise model of speech," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, USA, pp. 4609-4612, 2008.

[13] P. K. Lehana, R. Lochan, R. Lochan, and P. Abrol, "Enhancement of speech using Harmonic plus Noise Model," *IASTED, Signal and image processing*, Honolulu, Hawaii, USA, pp. 20-22, 2007.

[14] P. K. Lehana and P. C. Pandey, "Speech enhancement during analysis-synthesis by harmonic plus noise model," *Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 3039-3041, 2006.

[15] P. K. Lehana and P. C. Pandey, "Spectral mapping using multivariate polynomial modeling for voice conversion", Ph.D.Thesis, Department of Electrical Engineering, IIT Bombay, 2012.

[16] R. F. Chen, C. F. Chan, H. C. So, J. S. C. Lee, and C. Y. Leung, "Speech enhancement in car noise environment based on an analysis-synthesis approach using harmonic plus noise model," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4413-4416, 2009.

[17] R. Singh, A. Kumar., and P. K. Lehana, "Investigations of the Quality of Speech Imitated by Alexandrine Parrot (Psittacula~eupatria)," *Circuits, Systems, and Signal Processing*, vol. 36, no. 6, pp. 2292–2314, 2016.

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, "Genetic algorithm based speech enhancement in noisy environment," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.

ISSN: 2454-1532

Neha Jain, Sandeep Singh Gill, and Parveen Kumar Lehana, “Genetic algorithm based speech enhancement in noisy environment,” *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 3, pp. 27-34, 2018.