ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, Vol 8(33), DOI: 10.17485/ijst/2015/v8i33/77774, December 2015 ISSN (Online) : 0974-5645

Surround Noise Cancellation and Speech Enhancement using Sub Band Filtering and Spectral Subtraction

Vignesh Ganesan and Sangeetha Manoharan*

Department of Electronics and Communication Engineering, S. R. M. University, Kattankulathur, Kancheepuram District - 603203, Tamil Nadu, India; vignesh_thyagarajan@srmuniv.edu.in, sangeetha.m@ktr.srmuniv.ac.in

Abstract

Background/Objectives: To propose an algorithm based on Sub Band filtering, Spectral Subtraction, and Voice Activity Detector (VAD) for surround noise cancellation and speech enhancement. **Methods/Statistical Analysis**: Simulation study has been carried out to test the performance of the algorithm is tested for two real time speech signal (both male and female), considering three different surround noise sources such as fan, AC duct and Computer at various SNR dB levels. **Findings**: Simulation results show that the algorithm works efficiently in reducing the surround noise, which when mixed in a clean speech signal. For male speech signal corrupted by noise from fan, has -55 dB reduction of noise level, whereas for the female speech signal corrupted by noise from AC duct , the noise level is reduced to -65 dB and for Male speech signal corrupted by noise from computer, the noise level is reduce to -75 dB. **Conclusions/Improvements**: Presently the algorithm is proposed for stationary sound produced in a room, mainly for teleconferencing application, but the algorithm can also be tested for musical noise. Algorithm can be extended for sudden non stationary noises.

Keywords: Spectral Subtraction, Stereo, Sub Band Filtering, Surround Noise Cancellation, SQAM, VAD

1. Introduction

1.1 Overview

Noise defines any unwanted sound. Sounds, which are particularly loud, disturb people or make it difficult to here, wanted sounds are all noise. In communication aspect, some of the noises, which are common are: Cross talk, electrical noise, feedback etc. Acoustic noise is a type of noise that could be anything from a quiet but annoying to loud and harmful. These types of noise can be perceived either physiologically or psychologically. Psychological noise is perceived when our conscious awareness shifts its attention to the noise rather than letting it filter through our subconscious mind, where it goes unnoticed. Sound intensity follows an inverse square law, with distance from the source. Doubling the distance from noise source reduces its intensity by a factor of 4 dB or 6 dB.

Surround noise cancellation is one of the important aspects that are carried out in a teleconferencing system or hands free telephony. In a teleconference system, the noise sources could be from a computer, Air Conditioner (AC) duct, fan etc. In a hands free telephony inside a car, the noise source could be from car engine noise, outside vehicle and ambient wind. It is even possible that surround noise sources are captured at a higher level than the speech signal. In order to improve the speech signal quality, noise reduction algorithm is applied. Speech enhancement aims at improve the quality of the speech signal.

In[1], a system with Minimum Mean Square Estimator (MMSE) Short Time Spectral Amplitude (STSA) estimator is proposed for noise reduction. The MMSE STSA estimator is derived based on modeling speech and noise spectral components as statistically independent Gaussian random variables. The proposed approach results in a significant reduction of the noise and provides enhanced speech with colorless residual noise.

The concept of surround noise cancellation with channel decorrelation is used in Acoustic Echo Cancellers (AECs)[2,3]. Masato Akagi and Yoiti Suzuki in[4] proposed

* Author for correspondence

a two-microphone noise reduction method to deal with non-stationary interfering noises in multiple-noise-source environments in which the traditional two-microphone algorithms cannot function well. In the proposed algorithm, multiple interfering noise sources are regarded as one virtually integrated noise source in each sub-band, and the spectrum of the integrated noise is then estimated using its virtual direction of arrival. To do this, a direction finder for the integrated noise using only two microphones that performs well even in speech active periods was suggested. The noise spectrum estimate is further improved by integrating a single-channel noise estimation approach and then subtracted from that of the noisy signal, finally enhancing the desired target signal.

In[5], a new noise reduction scheme for hands-free telephony in a car environment was proposed which utilizes the time-domain Double Affine Projection (DAP) algorithm based on a sub-band structure. The DAP algorithm with high projection-order is suggested for operation in the low-frequency band, whereas a lower projection-order of the DAP algorithm can be applied to the high-frequency band. Simulation results suggest improved performance of the proposed noise reduction scheme in terms of noise attenuation and preservation of speech spectral components of the enhanced speech signals.

In[6], a comparative analysis of various speech enhancement methods in different noisy environments is presented. Speech signal corrupted with car noise and F16 noise is used for analysis. Spectral subtraction, iterative spectral subtraction, geometrical approach and Wiener filtering based methods, TSNR (Two Step Noise Reduction) and HRNR (Harmonic Regeneration Noise Reduction) methods are implemented in frequency domain for speech enhancement.

Anil Chokkarapu, Sarath C. Uppalapati and Abhiram Chintakuntla in[7] dealt with real-time implementation of spectral subtraction using Weighted OverLap Add (WOLA) filter bank to suppress noise. From the results, it can be analyzed that the noise is efficiently suppressed using spectral subtraction method and Power Spectral Density (PSD) of noise suppressed signal obtained from MATLAB and Digital Signal Processor (DSP) kit are studied and compared.

In[8], the use of Minimum-Variance Distortion less Response (MVDR) approach in single-channel speech enhancement in the short-time frequency domain is proposed. By applying optimal FIR filters to each sub-band signal, these filters reduce additive noise components with less speech distortion compared to conventional approaches. An algorithm to provide a blind estimation of temporal correlation of the speech signals based on a maximum-likelihood and maximum a-posteriori estimation is also proposed.

In speech processing, the short-time magnitude spectrum is believed to contain the majority of the intelligible information. The effect of the analysis of window duration on speech intelligibility in a systematic way is investigated[9]. A method for estimating the cross terms involving the phase differences between the noisy (and clean) signals and noise is proposed[10]. Analysis of the gain function of the proposed algorithm indicated that it possesses similar properties as the traditional MMSE algorithm. In[11], global optimal solution for active noise cancellation using Genetic algorithm technique is proposed.

In this paper, an algorithm for surround noise suppression is proposed that utilizes the concept of spectral subtraction. A compact VAD has been designed to identify the speech and the noise portion. The speech signal has been sub band filtered using sub band filtering process. This is done to reduce the computational complexity and also helps to work in real time with minimum delay. This algorithm also reconstructs the speech signal and hence the sound perception is almost unchanged.

1.2 Organization of the Paper

Section 2 discusses in detail the overview of sub band filtering, VAD and spectral subtraction. Section 3 presents the algorithm for surround noise cancellation and the data formulations. Section 4 discusses the simulation results. Finally, section 5 presents the conclusions.

2. Sub Band Filtering and VAD

Sub band filtering is the process of segmenting the signal into small frequency bands called sub bands. Sub band filtering is achieved by several stages of Band Pass Filters (BPF). An n stage BPF yields n sub bands. For every sub band, conversion from time domain to frequency domain called analysis and inverse sub band filtering called synthesis which converts frequency domain to time domain is necessary[12]. The advantage this method is its simplicity, but on the other hand the computational complexity increases by 10 times when processed in real time, thereby increases the delay.

In the proposed algorithm, Weighted Overlap Add Transform (WOAT) is used to reduce the surround noise.

WOAT is implemented in real time and the segmentation is done in time domain. VAD is used to identify the noise and speech portion. In a teleconference, the advantage is there will not be continuous speech signal throughout the call. There will be certain silence period of few milli seconds to seconds. In such scenario, VAD identifies the speech with noise segment and the noise alone segment. This helps in reducing the noise completely during noise alone segment and can reduce sufficient noise during the speech signal segment[6].

VAD is mainly used in speech coding and speech recognition. VAD is language independent, so one detector can be used for several languages and regions. The performance of VAD is evaluated based on graphical analysis and perceptual hearing. The parameters used to evaluate VAD are Front End Clipping (FEC) that is introduced in passing from noise to speech activity, Mid Speech Clipping (MSC) that is due to speech misclassified as noise, OVER where the noise is interpreted as speech due to the VAD flag remains active in passing from speech activity to noise and Noise Detected as Speech (NDS) where the noise is interpreted as speech within a silence period.

These parameters give an approximate measurement of the subjective effect. It is therefore important to carry out subjective tests to evaluate VAD. This kind of test would require certain number of people to listen and judge the recordings containing the processed signal of VAD.

3. Algorithm for Surround Cancellation

The process of surround noise cancellation and speech enhancement involves the design of an accurate VAD, proper analysis and synthesis stages of sub band filtering apart from spectral subtraction. The block diagram of the algorithm used for noise reduction is shown in Figure 1. The first stage is sub band filtering, which filters the given speech signal into smaller sub bands. Analysis stage performs the conversion of signal from time domain to frequency domain. The second stage is VAD, where the presence of speech and noise alone signal is identified. The output of VAD is given to spectral subtraction, where the magnitude spectrum of the signal is either subtracted or attenuated depending on the presence of speech or noise alone signal respectively. After the reduction of noise, the speech signal is reconstructed at the synthesis stage. The mathematical representation of the block outputs are as follows:

The noisy speech signal at the input of the sub band filtering process is given by

$$y(n) = x(n) + n(n) \quad (1)$$

where $x(n)$ is the noise free speech signal and $n(n)$ is the surround noise added to the speech signal. Sub band filtering converts time domain signal into frequency domain signal so Equation (1) is given as

$$Y(j\omega) = X(j\omega) + N(j\omega) \quad (2)$$

where $Y(j\omega)$, $X(j\omega)$, and $N(j\omega)$ are the fourier transform of the input noisy speech signal, noise free speech signal and noisy signal respectively. The output of the spectral subtraction is given as

$$X(j\omega) = Y(j\omega) - N(j\omega) \quad (3)$$

Based on properties like hass effect, cocktail party effect, few assumptions made for the algorithm and are as follows:

- The noise source is stationary and present throughout the conversation.
- Surround noises taken are Air Conditioner (AC) duct noise, computer noise and fan noise.
- During the conversation, the first 25 ms consist of only surround noise and no speech component present during that time. This duration was termed as Initial Silence (IS).
- The noise level is below -15dB.
- Number of Initial Silence frames calculated using Equation (4)

$$NIS = \frac{(IS \times f_s) - W}{(SP \times W_s) + 1} \quad (4)$$

Where IS-Initial Silence measured in ms, f_s is the Sampling Frequency, SP is the shift percentage and W is the Window Length .

Number of samples to be shifted is given as

$$NS = W \times SP \quad (5)$$

Total Number of frames (N) present can be calculated as

$$N = \frac{S - W}{NS} \quad (6)$$

The magnitude average is obtained by taking an average of previous, present and future speech signal samples and it is represented as

$$Y_{avg} = \frac{Y_{i-1} + Y_i + Y_{i+1}}{3} \quad (7)$$

Where Y_{i-1} is the previous speech signal sample, Y_i is the present speech signal sample and Y_{i+1} are the future speech signal sample. Using VAD, the presence of speech signal or noise alone signal is determined. If the result contains noise alone frame, then noise reduction is done by multiplying the noise alone frame with a constant α, whose value is varying from 0 to 0.99.

$$X = Y_n + \alpha \quad (8)$$

In a real time application such as Graphical User Interface (GUI) based applications, the value for alpha can be set as a varying unit.

If the result obtained from VAD signifies presence of speech, then the reduction of noise is carried out using Equation (9) and N is the mean of Initial Noise Power Spectrum.

$$D = Y - N \quad (9)$$

$$\text{where } D = \min\left[Y_{avg}^{i} - NY_{avg}^{i-1} - NY_{avg}^{i+1} - N\right] \quad (10)$$

The final stage is to reconstruct the speech signal. A mathematical equation for reconstructed speech signal size is obtained after several trail and errors with least possible error which will not be audible to human ear. This is the main advantage of using psycho-acoustic properties. The reconstructed speech signal size is defined as,

$$size = (N_F - 1) * ShiftLength\ h + W \quad (11)$$

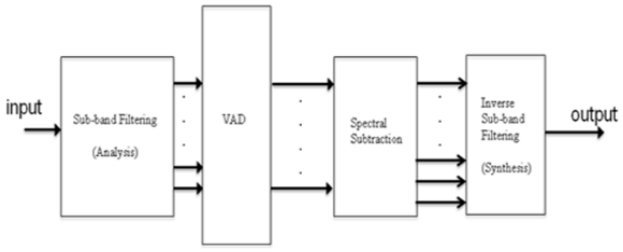

Figure 1. Block diagram of the algorithm.

4. Simulation Results

The algorithm for surround noise reduction is tested for both male and female speech signal for duration of 20 seconds taken from SQAM disk from ITU. These speech signals are sampled at the rate of 16 KHz. Three different noises say noise 1, 2 and 3 are originated from fan, AC duct and computer are recorded at -30dB, -40dB and -50dB respectively and mixed with the clean speech signal. Among these three noises, two noises are with male speech signal, and one noise is mixed with the female speech signal. This assumption has been taken to limit the number of simulation results which leads to redundancy. Practically all the combinations have been tested and their SNRs are compared and listed in the table.

Case 1: Male Speech Corrupted with Noise 1

In this case, the clean male speech signal is corrupted with noise 1 which is recorded at -30dB. Figure 2 shows the time domain representation of the input speech signal corrupted with noise 1. It is observed that the effect of noise in the signal is initially around – 30 dB.

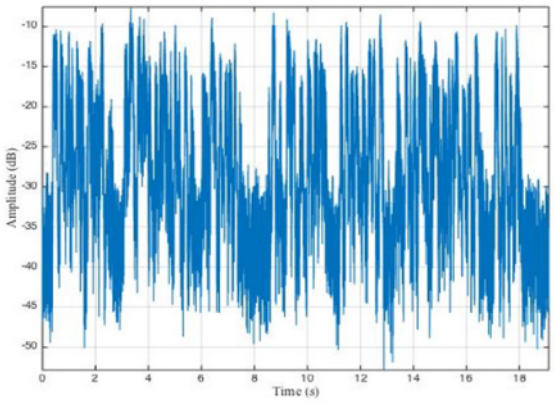

Figure 2. Time domain representation of the input noisy male speech signal.

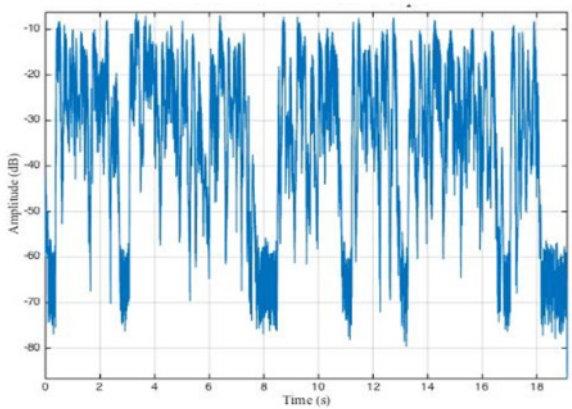

Figure 3. Time domain representation of the output male speech signal.

Figure 3 shows the time domain representation of the speech signal with noise being reduced. It is observed that the noise level is reduced to about -55dB. The magnitude of the speech signal is retained at around -10dB with a small fluctuation, from this we can analyze the level of the speech signal that has been retained and there will not be any amplitude variation for the listener.

Figure 4 shows the comparison of the input noisy speech signal and output noise free speech signal in time domain. The red mark signifies the presence of noise signal in the first subplot of Figure 4 which is completely removed in the second subplot of Figure 4 thus validating the proposed algorithm. This simulation result has been validated using listening tests.

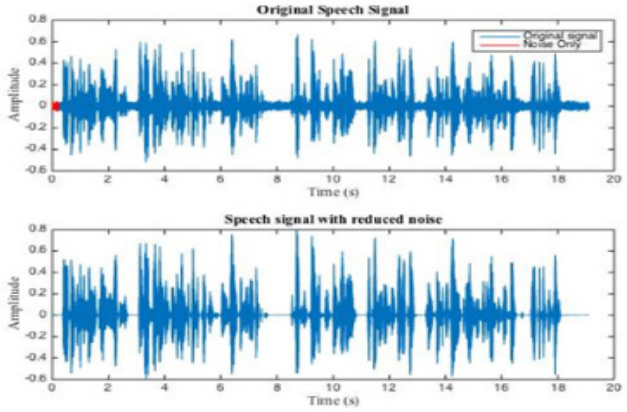

Figure 4. Time domain representation of input noisy male speech signal (subplot-1) and noise reduced output signal (subplot-2).

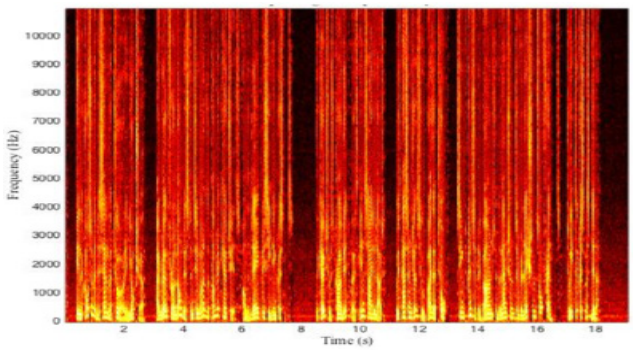

Figure 5. Spectrogram of input noisy male speech signal.

Figure 5 and Figure 6 shows the spectrogram of the input and output speech signals respectively. In Figure 5, the entire time duration is filled with color code of varying intensity because of the presence of noise. The white color indicates the highest intensity of the speech signal, followed by yellow being high, red being very intermediate intensity, and black being the least with almost zero intensity. The clean speech signal has pauses in between, which is considered as the silence segment of the speech signal. The effect of surround noise will be more during the silence period which is the black are of the Spectrogram result obtained in Figure 5.

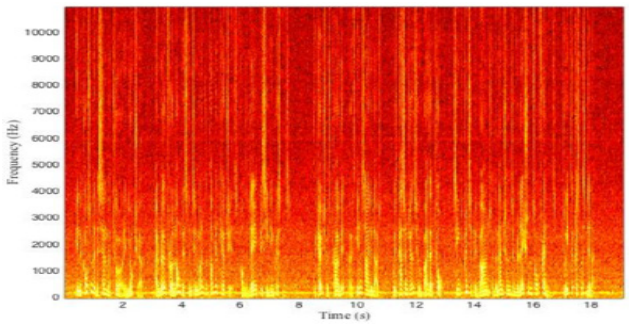

Figure 6. Spectrogram of output noise reduced male speech signal.

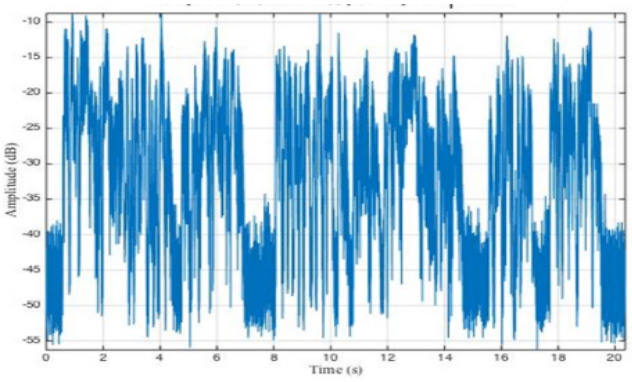

Figure 7. Time domain representation of the input noisy female speech signal.

The main objective of the proposed algorithm is to reduce the surround noise and enhance the speech signal which is visualized in Figure 6 which shows the spectrogram plot of the output noise reduced speech signal. In Figure 6, during the silence segments, there is still some intensity of noise present in the lower frequencies, but those noises are not audible and this can be verified with the levels plot in Figure 2 and Figure 3.

Case 2: Female Speech Corrupted with Noise 2

In this case, the clean female speech signal has been corrupted with noise 2 which was recorded at -40dB. Figure 7 show the time domain representation of the noise speech signal corrupted with noise 2. It can be observed that the effect of noise in the signal is initially around -40dB. In Figure 8 show the output speech signal with noise being reduced to -65dB.

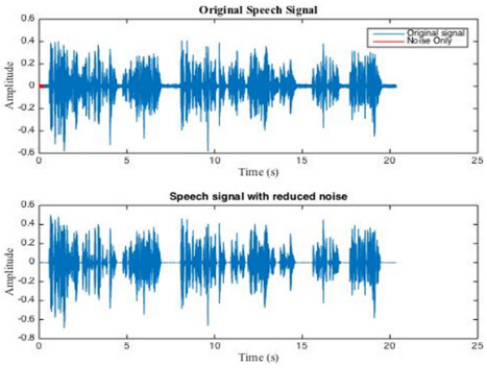

Figure 8. Time domain representation of the output noise reduced female speech signal.

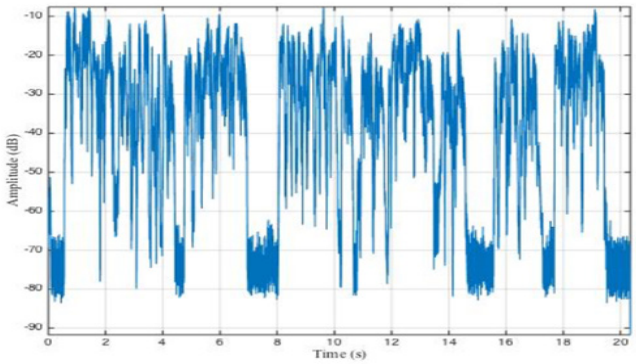

Figure 9. Time domain representation of input noisy female speech signal (first subplot) and noise reduced output signal (second subplot).

Figure 9 shows the comparison of the input noisy speech signal and output noise free speech signal in time domain. In the first subplot of Figure 9, the red portion indicates the presence of noise, and in the second subplot of Figure 9, it is observed that the noise is reduced completely. This simulation result has been validated using listening tests.

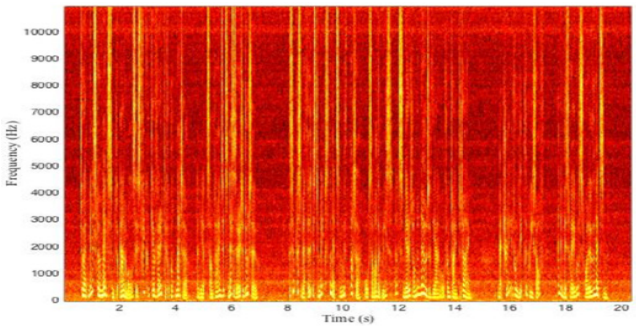

Figure 10. Spectrogram of input noisy female speech signal.

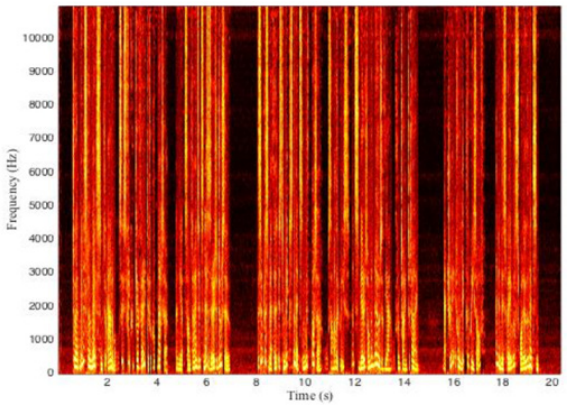

Figure 11. Spectrogram of output noise reduced female speech signal.

Figure 10 and Figure 11 show the spectrogram of the input noisy speech signal and the output noise free speech signal respectively. Red and yellow color code shows the two different intensity levels. During the silence period of the speech signal, the effect of surround of noise will be more. On comparing Figure 10 with Figure 11, the effect of noise is reduced completely. On comparing the results with respect to frequency axis, during the silence segments, there is only small portion of noise present in less than 1000Hz of frequency band compared to few thousands of Hz in the male speech signal.

Another interesting observation of the proposed algorithm is witnessed by comparing the spectrogram results obtained in Figure 5 and Figure 10, where there is more number of yellow lines in the female speech signal during the higher frequency range. This verifies that the female audio will have higher frequency content when compared to the male audio.

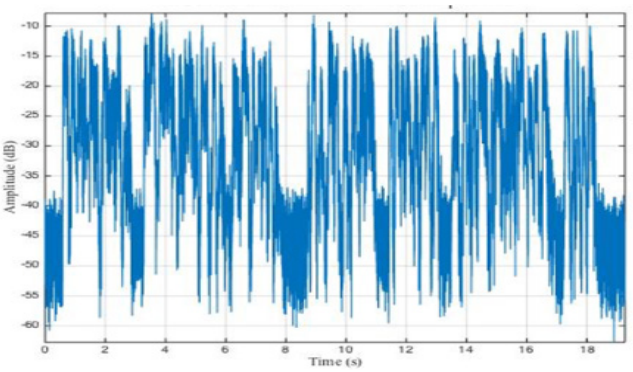

Figure 12. Time domain representation of the input noisy male speech signal with noise 3.

Case 3: Male Speech Corrupted with Noise 3

In this case, the clean male speech signal has been corrupted with noise 3, which was recorded at -50dB. The time domain representation of the input noisy male speech signal is shown in Figure 12 where the effect of noise in the speech signal is initially around -50dB. In Figure 13, the noise level alone is reduced to -75dB.

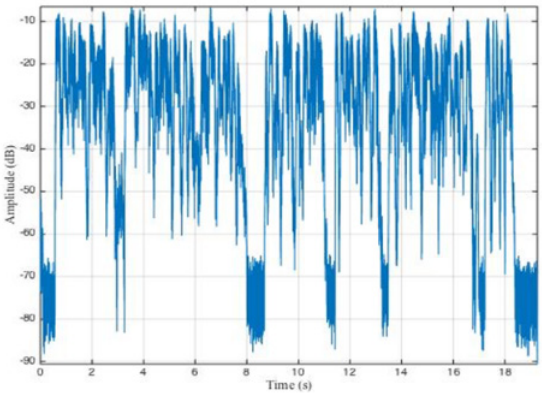

Figure 13. Time domain representation of the output noise reduced male speech signal with noise 3.

Figure 14 shows the comparison of input noisy speech signal and the output noise free speech signal in time domain. The red portion in subplot 1 of Figure 14 signifies the noise signal, and from the second subplot of Figure 14, it is observed that the noise signal has been reduced completely. This simulation result has been validated using listening tests. Figure 15 and Figure 16 shows the spectrogram plot of the input noisy speech signal and the output noise free speech signal.

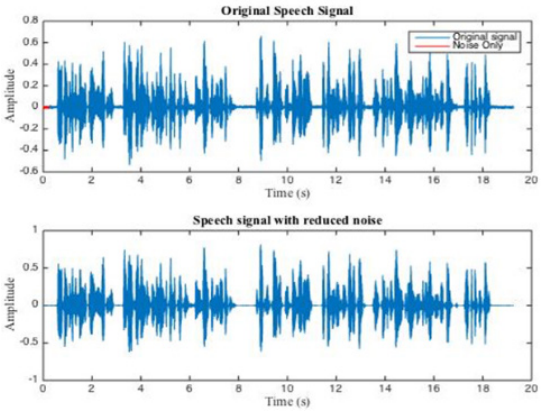

Figure 14. Time domain representation of the input noisy speech signal (subplot-1) and noise reduced output signal (subplot-2).

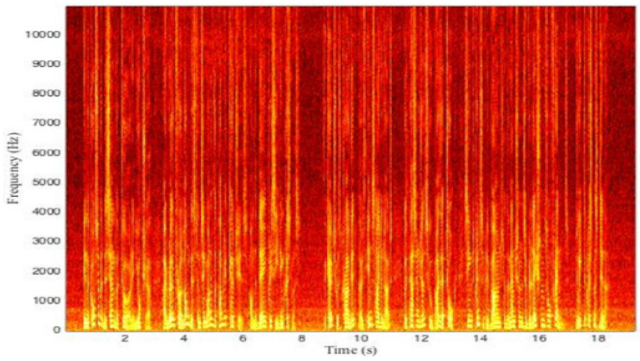

Figure 15. Spectrogram of input noisy male speech signal with noise 3.

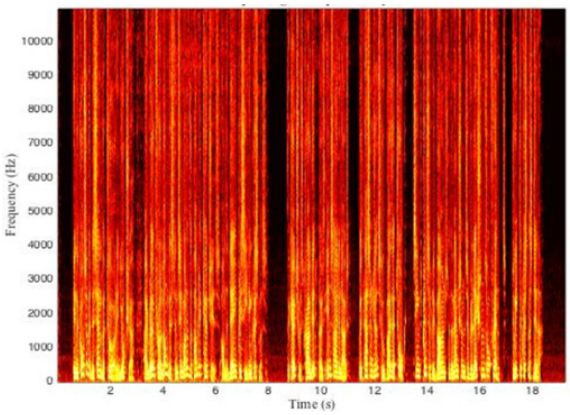

Figure 16. Spectrogram of output noise reduced speech signal.

4.1 Improvement in Signal to Noise Ratio (SNR) for Various Noise Signal Sources

Signal to Noise Ratio (SNRs) of the input speech signal, output speech signal and the improvement in SNR has been compared using a tabular form. Table 1 shows the comparison of SNR and noise level in dB for male speech signal. It is observed that, there is a significant improvement in the SNR level. Average improvement in SNR is around 19dB. The noise level has also been reduced drastically and because of this the noise is not audible at the final output speech signal.

Table 1. Comparison of SNR and noise level in dB for male speech signal

Male Speech Signal	I/P SNR (dB)	O/P SNR (dB)	Improvement in SNR (dB)	I/P Noise Level (dB)	O/P Noise Level (dB)
Noise 1	14.8	34.68	19.88	-30	-48
Noise 2	23.19	40.71	17.52	-40	-60
Noise 3	24.23	43.46	19.23	-50	-75

Table 2. Comparison of SNR and noise level in dB for female speech signal

Female Speech Signal	I/P SNR (dB)	O/P SNR (dB)	Improvement in SNR (dB)	I/P Noise Level (dB)	O/P Noise Level (dB)
Noise 1	13.34	26.50	13.16	-30	-48
Noise 2	23.19	40.71	17.52	-40	-60
Noise 3	23.61	37.94	14.33	-50	-75

Table 3. Comparison of I/P and O/P SNR in dB with noise level -20dB

Category of Speech signal	I/P SNR (dB)	O/P SNR (dB)	Improvement in SNR (dB)	I/P Noise Level (dB)	O/P Noise Level (dB)
Male	4.35	12.19	7.84	-20	-25
Female	-1.61	7.27	8.88	-20	-25

Table 2 contains the values regarding the case where female speech signal is mixed with various noises. It can be identified that there is a significant improvement in the SNR level. Average improvement in SNR is around 14dB. The output noise level is an average approximation and has been obtained using a levels plot. The above two cases of speech signals were mixed with three different noises. The simulation was carried out to test the efficiency of the algorithm when the noise level was greater than -20dB and the improvement in SNR is shown in Table 3.

From Table 3, it can be observed that the input and output SNR remains low, this proves that there is heavy influence of noise in the speech. The improvement in SNR is about 8dB. This makes the speech signal audible but along with the noise. The noise has not been reduced completely and it can be verified when comparing the input and output noise level.

5. Conclusions and Future Scope

The concept of surround noise cancellation is very vital in any teleconference system. In this paper, an algorithm for surround noise cancellation using sub band filtering and spatial subtraction is proposed. Simulation results show that the algorithm works efficiently in reducing the surround noise, which when mixed in a clean speech signal. This algorithm can also be extended for reducing noise in the low frequency range of less than 700 Hz. Currently this algorithm works for stationary sound. This could further be modified and extended to make it work for sudden non-stationary noise. This algorithm can also be tested for musical noise.

6. References

1. Ephraim Y, Malah D. Speech enhancement using a minimum mean square error short time spectral amplitude estimator. IEEE Transactions on Acoustic Speech and Signal Processing. 1984; 32(6):1109–21.
2. Vignesh G, Sangeetha M. Stereo channel decorreation using combined psychoacoustic approach. International Journal of Applied Engineering Search (IJAER). 2015; 6(10):156–65.
3. Benesty J, Morgan D, Sondhi MM. A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. IEEE Transactions on Speech Audio Processing. 1998; 6(2):156–65.
4. Akagi M, Suzuki Y. A two-microphone noise reduction method in highly non-stationary multiple-noise-source environments. International Journal on IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. 2008; 91(6):1337–46.
5. Thumchirdchupong H, Tangsangiumvisai N. Two-microphone noise reduction scheme for hands-free telephony in a car environment. Proceeding of the 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON); Krabi. 2013 May 15-17. p. 1–6.
6. Goel P, Saxena P, Chandra M, Gupta VK. Comparative analysis of speech enhancement methods. Proceedings of the 10th International Conference on Wireless and Optical Communications Networks (WOCN); Bhopal. 2013 July 26-28. p. 1–5.
7. Chokkarapu A, Uppalapati SC, Chintakuntla A. Implementation of spectral subtraction noise suppressor using DSP processor. International Journal of Computer Science and Telecommunications. 2013; 4(3):29–33.
8. Schasse A, Martin R. Estimation of subband speech correlations for noise reduction via MVDR processing. IEEE Transactions on Audio, Speech and Language Processing. 2014; 22(9):1355–65.
9. Paliwal K, Wo jcicki K. Effect of analysis window duration on speech intelligibility. IEEE Signal Processing Letter. 2008; 15(1):785–8.
10. Lu Y, Loizou PC. A geometric approach to spectral subtraction. International Journal on Speech Communication. 2008; 50(6):453–66.
11. Ahmed SF, Memon AR, Azim Ch F, Desa H. Global optimal solution for active noise control problem. Indian Journal of Science and Technology. 2011; 4(9):1015–20.
12. Boll S. Suppression of acoustic noise in speech using spectral subtraction. IEEE Transaction on Acoustic, Speech and Signal Processing. 1979; 27(2):113–20.