World Wide Journal of Multidisciplinary Research and Development

WWJMRD 2015; 2(8): 7-14
www.wwjmrd.com
Impact Factor MJIF: 4.25
e-ISSN: 2454-6615

Toyin Omotayo
Electrical/Electronics Engineering, University of Ilorin Ilorin, Kwara State, Nigeria

Lateef Afolabi
Electrical/Electronics Engineering, Federal Polytechnic Offa, Offa, Kwara State, Nigeria

Frederick Ehiagwina
Electrical/Electronics Engineering, Federal Polytechnic Offa, Offa, Kwara State, Nigeria

Titus Adewunmi
Electrical/Electronics Engineering, Federal Polytechnic Offa, Offa, Kwara State, Nigeria

Correspondence:
Frederick Ehiagwina
Electrical/Electronics Engineering, Federal Polytechnic Offa, Offa, Kwara State, Nigeria

Development of spectral subtraction algorithm for enhancement of noisy speech signal of electricity generator

Toyin Omotayo, Lateef Afolabi, Frederick Ehiagwina, Titus Adewunmi

Abstract
Speech enhancement entails a process of reducing noise and distortions by increasing the quality and intelligibility of a speech signal. This paper presents evaluation of spectral subtraction algorithm for noisy speech (samples taken in an environment where electricity generator is operated) without losing any part of the speech signal in terms of quality, quantity and without much computational and time complexity enhancement at different signal to noise ratios (SNR). Spectral subtraction was carried out on noisy speech samples at different SNR. The Noise removal algorithm was implemented using Matlab software. The corresponding spectrum was computed using the DFT (Discrete Fourier Transform) which removes the noise from the noisy speech and the corresponding spectrum was reconstructed in the time domain using the Inverse Discrete Fourier Transform (IDFT). The algorithms performance was evaluated by varying the Signal to Noise Ratio (SNR). The result indicates the optimal SNR values for electric generator noisy Speech Samples at -5dB, 5dB, 10Db, 15Db and 20dB. The spectral subtraction algorithms perform excellently in SNR range of -5.0000dB to 17.0500dB without any loss of part of the speech signal.

Keywords: IDFT, SNR, spectral-subtraction, speech-enhancement

Introduction

Enhancing the quality and intelligibility of noisy speech signal has attracted the interest of researchers over the years. The goal has been to improve quality, intelligibility and degree of listener fatigue of the speech signal [1]–[5]. Speech enhancement is an aspect of speech processing [6] used to manage the effects of noise. And it can be classified into, single channel, dual channel or multi-channel enhancement. Although the performance of multi-channel speech enhancement is better than that of single channel enhancement [7], the single channel speech enhancement is still a significant field of research interest because of its simple implementation and ease of computation. In single channel applications, only a single microphone is available and the characterization of noise statistics is extracted during the periods of pauses, which requires a stationary assumption of the background noise. Among speech enhancement techniques popular among researchers is the spectral subtraction techniques.

In spectral subtraction, noise spectrum is estimated at silence region at the start of the speech, with the assumption that noise will be stationary throughout the speech, and this is not true in practice [1]. The estimation of the spectral amplitude of the noise data is easier than estimation of both the amplitude and phase. In [8], it is revealed that the short-time spectral amplitude (STSA) is more important than the phase information for the quality and intelligibility of speech.

Based on the STSA estimation, the single channel enhancement technique can be divided into two classes. The first class attempts to estimate the short-time spectral magnitude of the speech by subtracting a noise estimate. The noise is estimated during speech pauses of the noisy speech[8]. The second class applies a spectral subtraction filter (SSF) to the noisy speech, so that the spectral amplitude of enhanced speech can be obtained. The design principle is to select appropriate parameters of the filter to minimize the difference between the enhanced speech and the clean speech.

These two classes belong to the family of spectral subtractive-type algorithms [9]–[12]. The spectral subtraction method of single channel speech enhancement is the most widely used conventional method for reducing additive noise. Many improvements are proposed to deal with the problems typically associated to spectral subtraction such as remnant broadband noise and narrow band tonal noise referred as musical noise [3], [4]. Other variants of spectral subtraction include multi-band spectral subtraction [13]–[15], Wiener filtering [16], iterative spectral subtraction [3], [4], [12], [17], and spectral subtraction based on perceptual properties [10], [11], [18].

A brief overview of spectral subtraction speech enhancement are presented in this section. A subband noise reduction scheme for two-microphone communication system was presented by [19]. The scheme uses a subband structure with different noise reduction arrangement for different frequency band. For the high frequency band, a modified cross-spectral subtraction is employed to eliminate the associated decorrelated noise spectral component, while for low frequency band both spectral subtraction method and a variable noise subtraction parameter are used. Bharti [1] presented an algorithm based on short term energy that continuously updates the noise spectrum whether stationary or non-stationary. Oversampled DFT modulated filter bank was used to decompose the time series of speech into subbands of equal space by [20]. It capitalizes on the fact that most environmental noises do not influence speech spectrum uniformly with different frequency [14]. Meanwhile, [14] used weighted recursive averaging methodology to approximate noise power spectrum and later exerted subtraction of multi-band on the speech signal affected by noise. Also auditory masking threshold was computed, and subsequent associated subtraction factor was adjusted.

Deepak [21] exploited speech production features like glottal closure instants in time domain and vocal tract information in spectral domain. The desired speaker speech is perceptually enhanced using auditory perceptive feature in mel-frequency domain using mel-cepstral coefficients and its inversion using mel-log spectrum approximation filter. A means for removing remnant noise was proposed by [12], in which the output of multi-band spectral subtraction (MBSS) method is fed to the input again for next iteration process. After the MBSS method, the additive noise is changed to remnant noise, which is continuously re-calculated at each iteration. Meanwhile, [22] took into consideration cross-term containing noise spectral and speech in developing a modified spectral subtraction technique. Pink noise, white noise, and so on was collected from database by [23], who subsequently used non-linear spectral subtraction and MBSS technique for speech enhancement. Furthermore, [24] proposes a noise compensation carried out on time magnitude spectral following the geometric for speech enhancement.

Other works in this regard include that of [18], [25]–[30]. For instance, [25] estimated spectral subtraction mask using the frequency dependent reverberation time and decay rate to develop a dereverberation algorithm. A discussion of both the conventional spectral subtraction and a modified spectral subtraction for improving speech quality degraded by additive background noise and dynamic additive noise respectively was carried out by [26]. The authors in [27] compared results from audio-visual speech recognition with that of log-spectral minimum mean square error, and multi-band spectral subtraction methods for reducing additive noise in audio signal arte used. However, [18] have reservations about spectral subtraction and conventional speech enhancement techniques because they are not sufficient in describing highly non-stationary noise. Hence, in this paper, a simulation study of different forms of spectral subtractive-type algorithms is described. In addition, the power spectrum of the noise is estimated based on the first few milli-seconds of the noisy speech signal. This short duration is called the silence region, and it is useless just like the noise in some application of speech technology. Conventional Voice Activity Detection (VAD) algorithms are employed to monitor and detect these regions to ensure that a clean speech devoid of any blank or noise signal is obtained at the system output. However, the problem is that different noise has different effects on speech signals [9]. Thus, there is a need for an improved noise reduction algorithm effective for most noisy backgrounds at specified SNR. This paper proposes an improved spectral subtraction algorithm which takes cognizance of the real time occupancy noise of the noise speech signals and hence makes it suitable for different noisy environments.

Modeling of Noisy Speech Signal and Spectral Subtraction Filtering Technique

In spectral subtraction, input signal is segmented into frames and multiplied with hamming window. Discrete Fourier Transform (DFT) of these frames and separate magnitude and phase from speech is obtained. The noise power estimation and calculation of spectral weighting takes place on magnitude. Once the noise estimate is subtracted from speech spectrum magnitude, it is recombined with the original phase of the noisy signal. Inverse Fourier Transform (IDFT) is taken and output is obtained by overlap and add method.

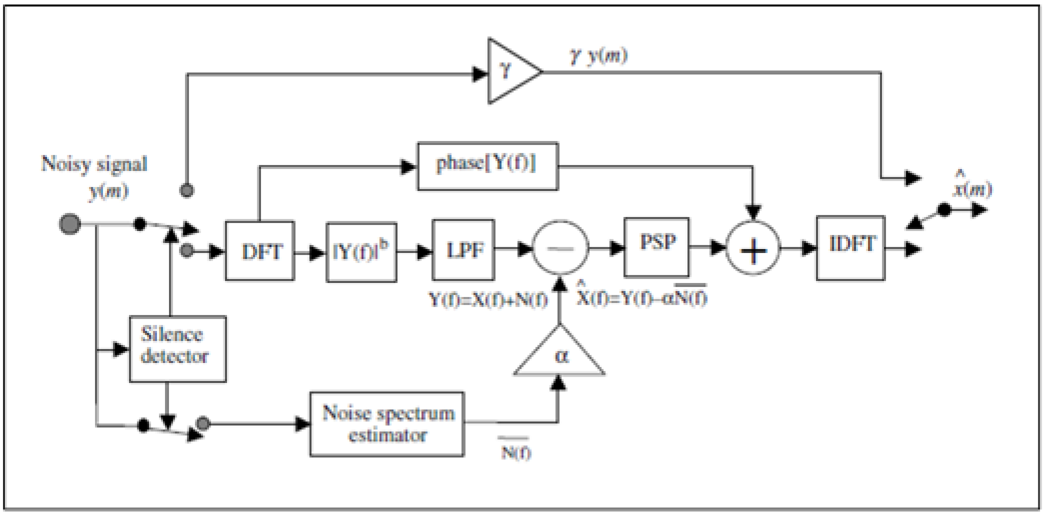

Fig. 1: Block Diagram of the General representation of Spectral Subtraction [8]

From fig.1, the noisy speech signal *y (m)* which comprises of both the clean and noise signal is framed and windowed, each frame which is 20ms in length is simultaneously processed by both the VAD and DFT blocks, the voice activity detection (VAD) block is a binary classifier, that is the output can either be 1 or 0, the output is 1 if the frame contains speech samples and 0 otherwise. The DFT block converts each frame to its frequency domain to obtain the spectral magnitudes and phase. The non-speech segments or frames from the VAD block are used to determine the noise estimate, the noise estimate or magnitude from the noise estimation block is subtracted from the spectral magnitudes. The difference is then combined with the phase obtained from the DFT block to generate the enhanced speech spectral magnitudes, the output magnitude are then processed by the IDFT to obtain the enhanced speech signal in time domain. Speech which is corrupted by noise can be expressed as equation (1). If $y(m)$is speech with noise, $x(m)$ is clean speech signal and $n(m)$ is noise process, all in the discrete time domain, then,

$$y(m) = x(m) + n(m) \qquad (1)$$

Consequently, $x(m)$is estimated from $y(m)$.

If the noise process is represented by its power spectrum estimate $|\tilde{N}(w)|^2$, that of the noisy speech is $|G(w)|^2$, the power spectrum of the clean speech estimate $|X(w)|^2$ can be written as equation (2):

$$|\hat{X}(w)|^2 = |Y(w)|^2 - |\bar{N}(w)|^2 \qquad (2)$$

Since the power spectrum of two uncorrelated signals is additive. The clean speech phase $\theta Y(w)$ is estimated directly from the noisy speech signal phase $\theta G(w)$ as in equation (3).

$$\theta X(w) = \theta Y(w) \qquad (3)$$

Thus a general form of the estimated speech in frequency domain can be written as in equation (4).

$$\hat{X}(w) = \left(\max(|Y(w)|^2 - k|\bar{N}(w)^2 \;\; ,0)^{\frac{1}{2}}\right) . e^{j\theta Y(w)} \qquad (4)$$

Where, k > 1 is used to overestimate the noise to account for the variance in the noise estimate. The inner term $|Y(w)|^2 - k|\tilde{N}(w)|^2$ is limited to positive values, since it is possible for the overestimated noise to be greater than the current signal.

The environment where the speech collected was influenced by electricity generator.

Results and Discussion

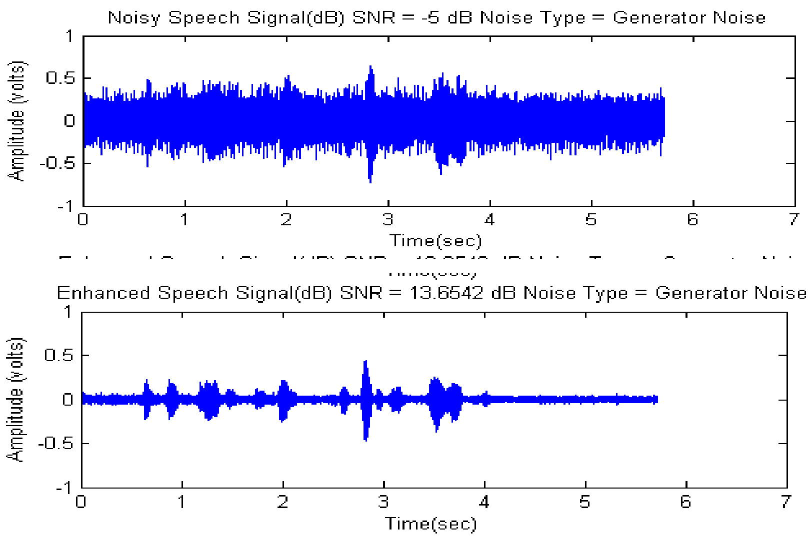

Fig. 2: The generator noisy speech signal at signal to noise ratios of -5dB in time domain

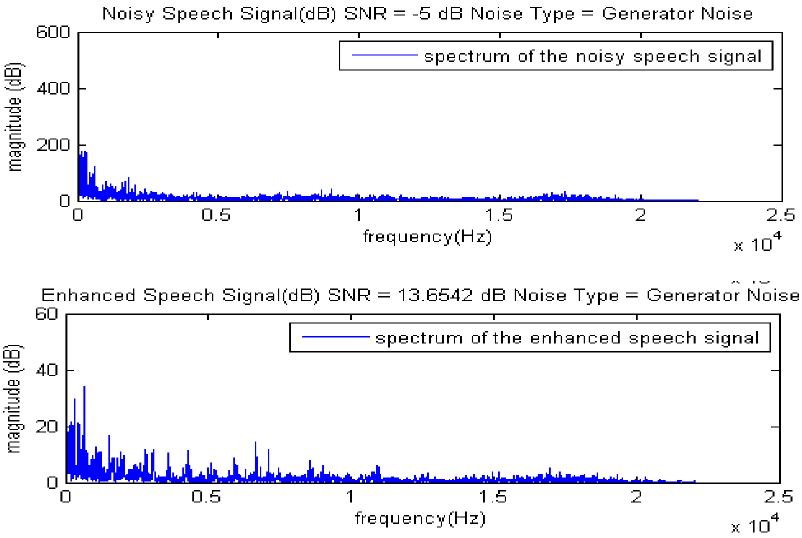

Fig. 3: The generator noisy speech signal at signal to noise ratios of -5dB in frequency domain

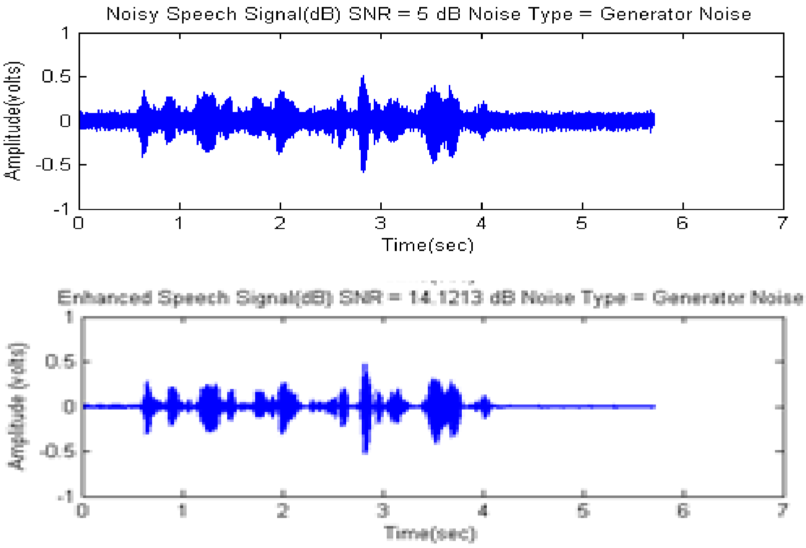

Fig. 4: The generator noisy speech signal at signal to noise ratios of 5dB in time domain

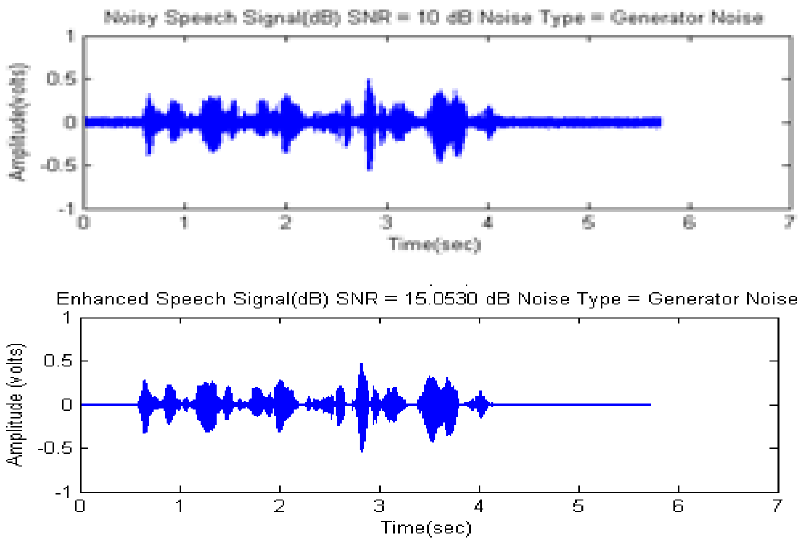

Fig. 6: The generator noisy speech signal at signal to noise ratios of 10dB in time domain

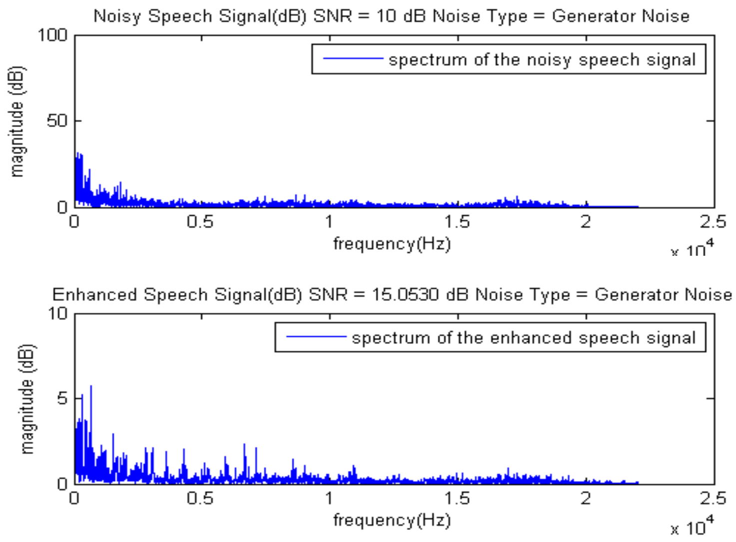

Fig. 7: The generator noisy speech signal at signal to noise ratios of 10dB in frequency domain

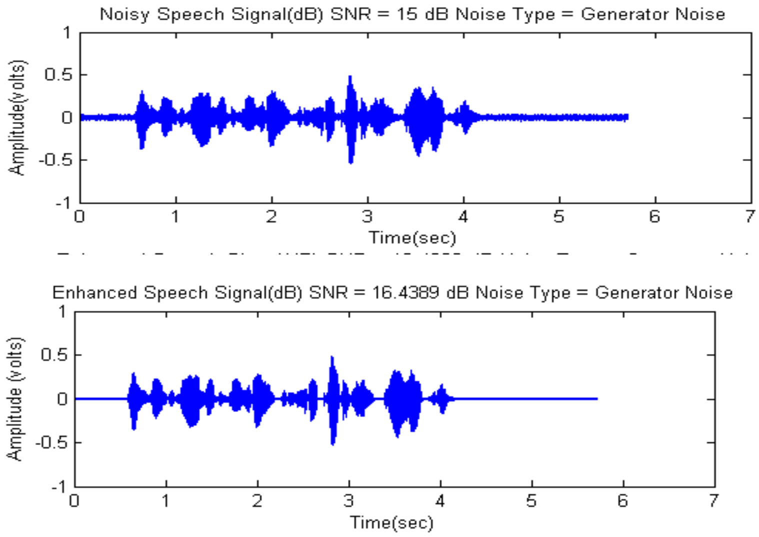

Fig. 8: The generator noisy speech signal at signal to noise ratios of 15dB in time domain

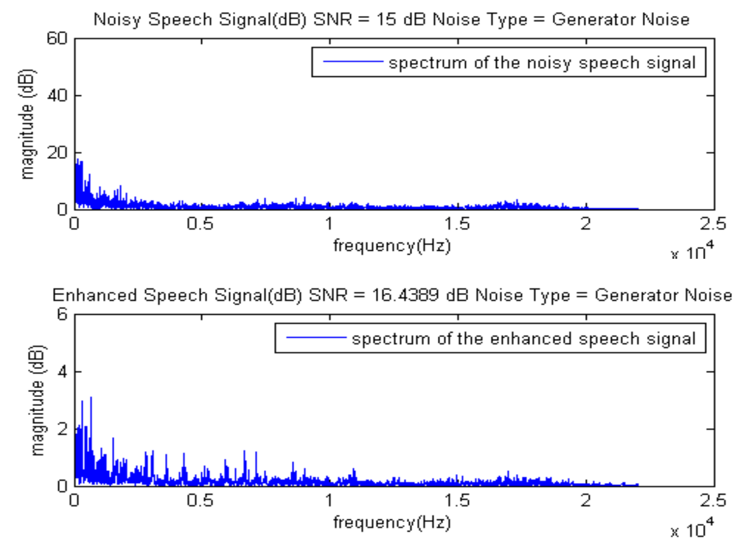

Fig. 9: The generator noisy speech signal at signal to noise ratios of 15dB in frequency domain

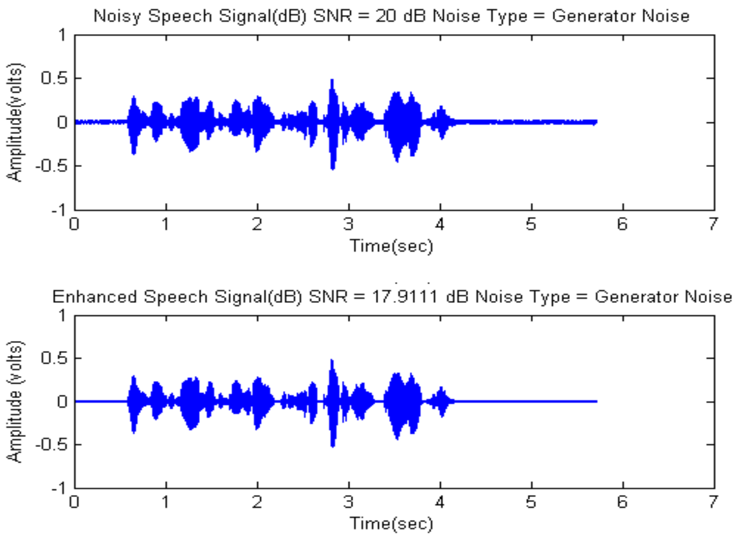

Fig. 10: The generator noisy speech signal at signal to noise ratios of 20dB in time domain

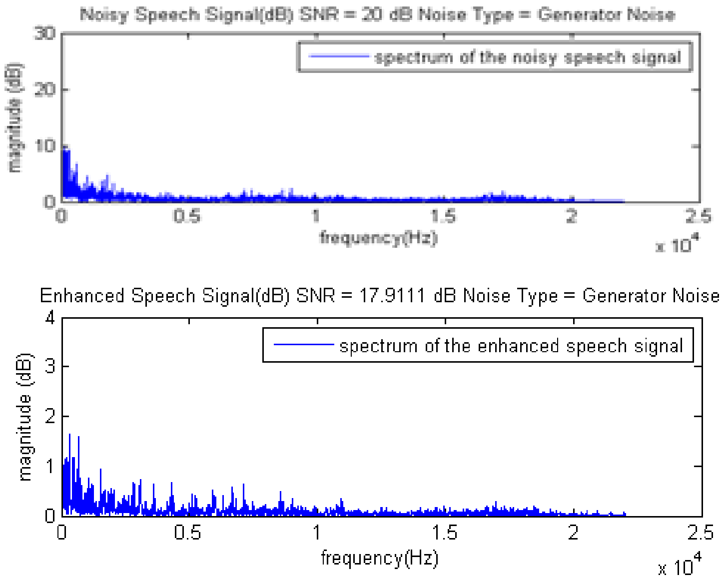

Fig. 11: The generator noisy speech signal at signal to noise ratios of 20dB in frequency domain

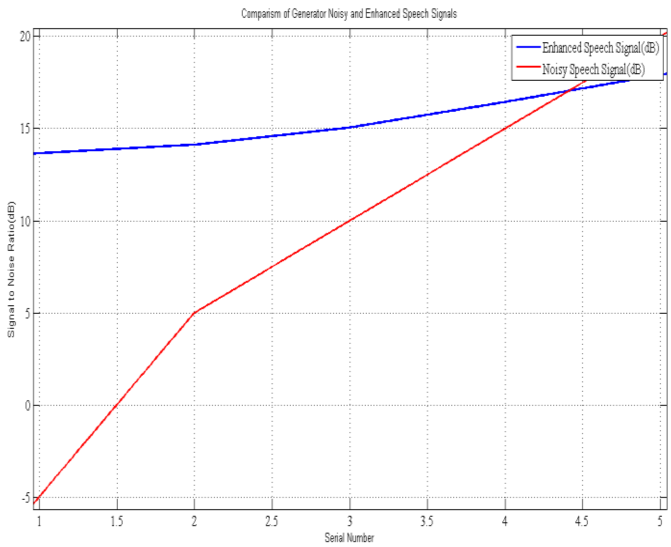

Fig. 12: Comparison of both the Generator Noisy and Enhanced Speech signals

The results are shown in figs. 1-11. At -5dB, 5dB, 10dB, and 15dB there exist a tremendous improvement in the signal to noise ratio, and the result are 13.7629dB, 14.5620dB, 15.6777dB, and 17.1748dB respectively. It was noticed that at 20dB and above there was deterioration in the signal to noise ratio i.e. at 20dB the SNR result value gives 18.6374dB. From Fig. 12, a general graph that reveals the signal to noise ratios for generator noisy speech. Beyond the upper limit point there can be no further enhancement to the noisy speech signals instead it deteriorates. The upper limit point is 17.0500dB, Therefore any SNR value above 17.0500dB will result in the deterioration of the generator noisy speech signals.

Conclusion

It can be concluded that the Spectral Subtraction algorithm is efficient, fast and reliable for the generator noisy speech signal. The result shows that spectral subtraction noise reduction technique has a limit of -5.0000dB to 17.0500dB SNR value for generator noisy speech signal. Any SNR above 17.05dB will results in loss of part of the signal.

References

1. S. S. Bharti, M. Gupta, and S. Agarwal, “A new spectral subtraction method for speech enhancement using adaptive noise estimation,” in *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, 2016, pp. 128–132.
2. L. Sui, X. Zhang, J. Huang, G. Zhao, and Y. Yang, “Speech enhancement based on sparse nonnegative matrix factorization with priors,” *Systems and Informatics (ICSAI), 2012 International Conference on*. pp. 274–278, 2012.
3. R. Miyazaki, H. Saruwatari, T. Inoue, K. Shikano, and K. Kondo, “Musical-noise-free speech enhancement: Theory and evaluation,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 4565–4568.
4. R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano, and K. Kondo, “Musical-Noise-Free Speech Enhancement Based on Optimized Iterative Spectral Subtraction,” 2012.
5. T. Petsatodis, C. Boukis, F. Talantzis, and L. Polymenakos, “Cascaded dynamic noise reduction utilizing VAD to improve residual suppression,” *Digital Signal Processing (DSP), 2013 18th International Conference on*. pp. 1–6, 2013.
6. C. Schuler, M; Chugani, *Digital signal processing: a hand on approach*, Internatio. Mcgraw Hills companies, 2000.
7. W. Ephraim, Y.; Ari, H. L.; Roberts, *A brief survey of speech enhancement: the electrical engineering handbook*, 3rd ed. Boca Raton, FL, 2006.
8. P. C. Loizou, *Speech enhancement: theory and practice*, 1st ed. Taylor and Francis, 2007.
9. N. Upadhyay and A. Karmakar, “The spectral subtractive-type algorithms for enhancing speech in noisy environments,” *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*. pp. 841–847, 2012.
10. N. Upadhyay and A. Karmakar, “A Perceptually Motivated Multi-Band Spectral Subtraction Algorithm for Enhancement of Degraded Speech,” *Computer and Communication Technology (ICCCT), 2012 Third International Conference on*. pp. 340–345, 2012.
11. N. Upadhyay and A. Karmakar, “An auditory perception based improved multi-band spectral subtraction algorithm for enhancement of speech degraded by non-stationary noises,” *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*. pp. 1–7, 2012.
12. N. Upadhyay and A. Karmakar, “Single channel speech enhancement utilizing iterative processing of multi-band spectral subtraction algorithm,” *Power, Control and Embedded Systems (ICPCES), 2012 2nd International Conference on*. pp. 1–6, 2012.
13. P. Kamath, S.; Loizou, “A Multi-Band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise,” in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2002, pp. 4160–4164.
14. L. Cao, T. q. Zhang, H. x. Gao, and C. Yi, “Multi-band spectral subtraction method combined with auditory masking properties for speech enhancement,” in *Image and Signal Processing (CISP), 2012 5th International Congress on*, 2012, pp. 72–76.
15. A. Corporation, “White Paper Developing Multipoint Touch Screens and Panels With CPLDs Cursor Right click Left click,” *Baseline*, no. February, pp. 1–5, 2009.
16. F. E. Abd El-Fattah, A.; Dessouky, M. I.; Diab, S. M.; Abd El-samie, “Speech enhancement using an adaptive Wiener filtering approach,” *Electromagn. Res. M.*, vol. 4, pp. 167–184, 2008.
17. F. Polyt, O. Onabanjo, and O. Sta, “Numerical Approxima ation of Ge neralised Riccati Equ uations by Iterative Decomp position Me ethod,” vol. 9, no. 6, pp. 310–312, 2011.
18. S. P. Patil and J. N. Gowdy, “Exploiting the baseband phase structure of the voiced speech for speech enhancement,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6092–6096.
19. T. T. Aung, H. Thumchirdchupong, N. Tangsangiumvisai, and A. Nishihara, “Two-microphone subband noise reduction scheme with a new noise subtraction parameter for speech quality enhancement,” *IET Signal Processing*, vol. 9, no. 2. pp. 130–142, 2015.
20. Y. Cai and C. Hou, “Subband spectral-subtraction speech enhancement based on the DFT modulated filter banks,” in *Signal Processing (ICSP), 2012 IEEE 11th International Conference on*, 2012, vol. 1, pp. 571–574.
21. K. T. Deepak and S. R. M. Prasanna, “Foreground Speech Segmentation and Enhancement Using Glottal Closure Instants and Mel Cepstral Coefficients,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 7, pp. 1204–1218, 2016.
22. M. T. Islam, C. Shahnaz, and S. A. Fattah, “Speech enhancement based on a modified spectral subtraction method,” *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*. pp. 1085–1088, 2014.
23. N. G., Prabhakaran, G.; Indra, J.; Kasthuri, “Tamil speech enhancement using non-linear spectral subtraction,” in *Communications and Signal Processing (ICCSP), 2014 International Conference on*, 2014, pp. 1482–1485.

24. M. T. Islam, A. B. Hussain, K. T. Shahid, U. Saha, and C. Shahnaz, "Speech enhancement based on noise compensated magnitude spectrum," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*, 2014, pp. 1–5.
25. T. Jan and W. Wang, "Frequency dependent statistical model for the suppression of late reverberations," in *Sensor Signal Processing for Defence (SSPD 2012)*, 2012, pp. 1–5.
26. S. S. Meher and T. Ananthakrishna, "Dynamic spectral subtraction on AWGN speech," *Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on*. pp. 92–97, 2015.
27. K. Palecek and J. Chaloupka, "Audio-visual speech recognition in noisy audio environments," in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*, 2013, pp. 484–487.
28. S. K. Waddi, P. C. Pandey, and N. Tiwari, "Speech enhancement using spectral subtraction and cascaded-median based noise estimation for hearing impaired listeners," in *Communications (NCC), 2013 National Conference on*, 2013, pp. 1–5.
29. N. Tiwari, S. K. Waddi, and P. C. Pandey, "Speech enhancement and multi-band frequency compression for suppression of noise and intraspeech spectral masking in hearing aids," *2013 Annual IEEE India Conference (INDICON)*. pp. 1–6, 2013.
30. N. Tiwari and P. C. Pandey, "Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners," *Communications (NCC), 2015 Twenty First National Conference on*. pp. 1–6, 2015.