American Journal of Circuits, Systems and Signal Processing
Vol. 1, No. 2, 2015, pp. 47-55
http://www.aiscience.org/journal/ajcssp

Communication Systems Noise Reduction Based on Adaptive Spectral Subtraction Method

Isiaka A. Alimi*, Tusin D. Ebinowen

Department of Electrical and Electronics Engineering, School of Engineering and Engineering Technology, Federal University of Technology, Akure, Nigeria

Abstract

The spectral subtraction (SS) method is a well-known signal enhancement technique that reduces the effect of noise in a noisy signal in order to improve the signal quality. The SS works on the principle that noise spectrum estimate over the entire speech spectrum can be subtracted from the noisy signal. However, noise does not affect the speech signal uniformly over the entire spectrum at different frequency bands. Therefore, most implementations of the basic technique lead to anomaly known as "musical" tones artifacts in the enhanced signal. The abnormality can then be perceived as residual noise and speech distortion in the resulting signal. In this paper, we propose a multi-band spectral subtraction (MBSS) method using novel noise element suppression (NES). The proposed scheme gives comparatively better performance and the computation required is minimal. Furthermore, simulation results show that the proposed algorithm removes noise without removing the relatively low amplitude signal over the entire speech spectrum.

Keywords

Speech Enhancement, Musical Noise, Spectral Subtraction, Noise Element Suppression, Multi-Band, Sub-Band

Received: May 2, 2015 / Accepted: May 17, 2015 / Published online: June 18, 2015

1. Introduction

In recent years, voice communication technologies have been rapidly growing since the advent of mobile telephony systems in the commercial market. As the technologies evolve, new services are also introduced like teleconference systems and voice over internet protocol (VOIP). Noise suppression techniques are essential for the systems to operate efficiently and effectively because, the presence of noise often result in erroneous and unreliable communication systems [1]. Consequently, a method that can suppress the noise while maintaining the required sound quality is essential [2]. Boll proposed SS technique for suppressing the effect of noise acoustically added to the speech signals [3]. The approach is popular because of its simplicity and flexibility in concept and effectiveness in enhancing speech degraded by additive noise [4].

To implement SS, spectral magnitude of the received noisy signal is estimated and that of the noise spectrum is estimated from regions that are analyzed as "noise-only" using voice activity detection (VAD). The magnitude spectrum of noise is then subtracted from that of the noisy signal. The approach works under the assumption that noise signal is uncorrelated and remains relatively constant prior to and during voice activity. This implementation noticeably gives quality speech signal but generates musical noise through linear subtraction of noise across the entire speech spectrum [5], [6].

Current researches focus on a nonlinear subtraction process which is justified by the variation of signal-to-noise ratio across the speech spectrum. Also, the nonlinear approach shows that noise signal does not affect the speech signal uniformly over the whole spectrum because certain frequencies are affected more adversely than others [2], [7]. To prevent the variation of signal-to-noise ratio across the

* Corresponding author
E-mail address: compeasywalus2@yahoo.com (I. A. Alimi)

enhanced speech spectrum as well as destructive subtraction of speech while removing most of the residual noise, it is necessary to develop an appropriate scheme that will subtract only the necessary amount of noise spectrum from each frequency bin. The criteria for quantifying the amount of generated musical noise are proposed in [8] and [9].

In this paper, we propose a multi-band approach which is based on SS technique. The proposed approach maintains high speech quality and mitigates anomalies which are related to the linear subtraction technique by employing the noise element suppression (NES). The scheme divides the spectrum into frequency sub-band based on the nonlinear multiband frame. For each sub-band, the noise corrupted speech power in the past and the present time frames are compared to the statistics of the noise power in order to improve the determination of voice activity in the sub-band. In the subtraction process, a larger proportion of noise is removed from sub-band in which the voice activity is zero. Furthermore, in sub-band that contain speech, an NES is developed which allows for the removal of less noise during relatively low amplitude speech and more noise during relatively high amplitude speech.

The rest of the paper is organized as follows: in Section 2, mathematical description of the multi-band spectral subtraction (MBSS) and the associated techniques are discussed. Section 3 presents the methodology of the work. Section 4 contains the simulation results and analysis. Conclusions are drawn in Section 5.

2. Multi-Band Spectral Subtraction (MBSS)

Assume that a wanted signal $s(n)$ is corrupted by an additive noise $d(n)$, the received signal $r(n)$ is the sum of $s(n)$ and $d(n)$. The noisy signal model in the time domain is represented as [10]:

$$r(n) = s(n) + d(n) \tag{1}$$

where n is the discrete time index.

The application of SS relies on the fact that the estimated noise spectrum, $\widehat{D}(\omega)$ can be subtracted from the received signal spectrum, $R(\omega)$. The estimated output spectrum, $\hat{S}(\omega)$ can be transformed by the inverse fast Fourier transform (IFFT) to $\hat{s}(n)$.

The Fourier transforms of equation (1) in the frequency domain is:

$$R(\omega) = S(\omega) + D(\omega) \tag{2}$$

where $R(\omega)$, $S(\omega)$ and $D(\omega)$ are the Fourier transforms of the noisy signal $r(n)$, the wanted signal $s(n)$ and the noise signal $d(n)$ respectively and ω is the frequency variable.

The received signal $r(n)$ is buffered and divided into segments that consist of N samples length. Each segment is Hamming windowed and then transformed via discrete Fourier transform (DFT) to N spectral samples. Windowing has the advantage of alleviating the effects of discontinuities at the endpoints of each segment and also suppresses glitches; therefore, it avoids the broadening of the frequency spectrum caused by the glitches [7], [10].

In frequency domain, the windowing operation can be represented by:

$$R_w(\omega) = S_w(\omega) + D_w(\omega) \tag{3}$$

The equation describing SS may be expressed as:

$$\left|\hat{S}(\omega)\right|^p = |R(\omega)|^p - \alpha\left|\overline{D(\omega)}\right|^p \tag{4}$$

where $\left|\hat{S}(\omega)\right|^p$is an estimate of the original signal spectrum and $\left|\overline{D(\omega)}\right|^p$ is the time-averaged noise spectra. It is assumed that the noise is a wide-sense stationary random process.

For magnitude SS, the exponent p=1, and for power SS, p=2. The parameter α is for controlling the amount of noise subtracted from the noisy signal. For full noise subtraction, α =1 and for over-subtraction α >1.

A novel NES, α_k is proposed that can adaptively control the level of the noise to be subtracted from each

k^{th} sub-band. The magnitude SS of the estimated signal is given as:

$$\left|\hat{S}(\omega)\right| = |R(\omega)| - \alpha\left|\overline{D(\omega)}\right| \tag{5}$$

A multi-band approach to SS is proposed since noise has non-uniform spectral distribution and its equivalent band consists of K parallel subbands [10]. Furthermore, the signal spectrum is divided into K non-overlapping sub-band and the segmentations make it easy to performed SS independently across multiple frequency sub-band. Consequently, the estimate of the signal spectrum in the k^{th} channel is obtained by:

$$\left|\hat{S}_k(\omega)\right| = |R_k(\omega)| - \alpha_k\left|\overline{D_k(\omega)}\right| \quad 0 \le k \le K-1 \tag{6}$$

Evaluating the expectation of Equation (6), implies:

$$E\left[\left|\hat{S}_k(\omega)\right|\right] = E[|R_k(\omega)|] - \alpha_k E\left[\overline{|D_k(\omega)|}\right] \tag{7}$$

$$= E[|S_k(\omega)|] + E\left[\overline{|D_k(\omega)|}\right] - \alpha_k E\left[\overline{|D_k(\omega)|}\right]\Big]_{E[|R_k(\omega)|]=E[|S_k(\omega)|]+E[|D_k(\omega)|]} \tag{8}$$

$$\approx E[|S_k(\omega)|] \tag{9}$$

For signal restoration, the magnitude estimate of the required signal, $|\hat{S}_k(\omega)|$ is combined with the phase of the noisy signal and then inverse Fourier transformed. The resulting signal is overlap added to reconstruct the enhanced output sequence. The block diagram of the proposed NES technique is depicted in Fig. 1.

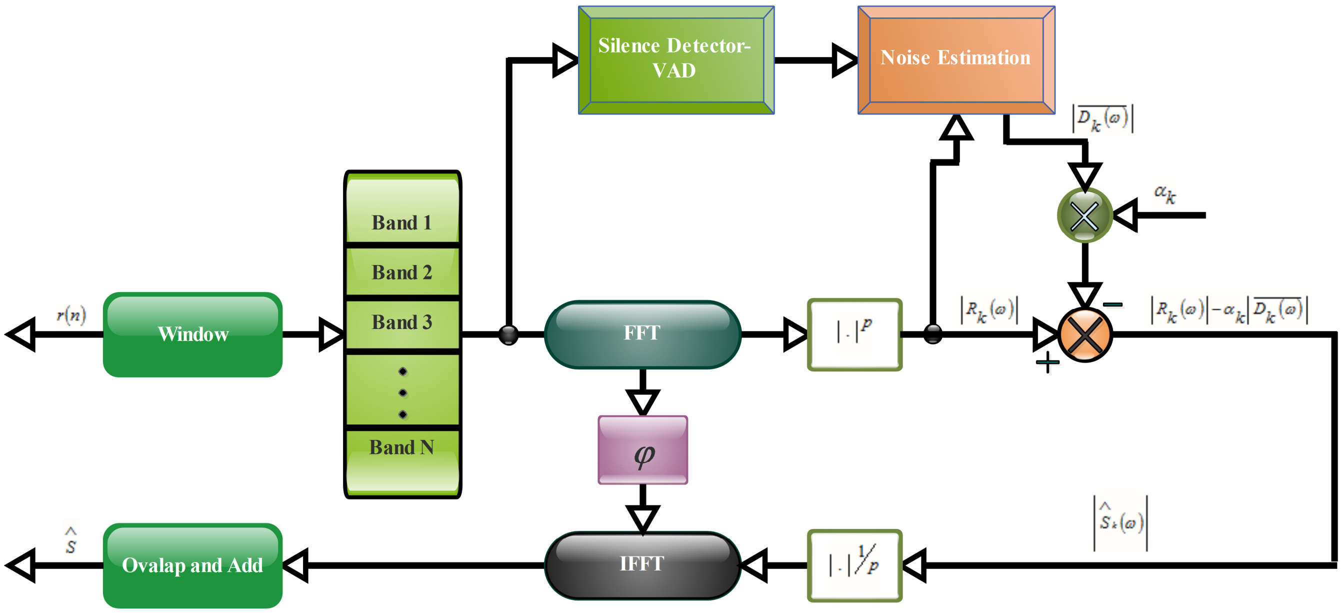

Fig. 1. Block Diagram of the Proposed NES Technique.

3. Methodology

In this work, three sinusoidal signals are simulated at 8 kHz sampling frequency and their spectra are determined. The three simulated sinusoidal signals are combined and the resulting composite spectrum is determined. Furthermore, 1024 samples of simulated noise signal are added to the composite signal to produce the noisy signal. The noisy signal is then windowed using a 20 ms (160 samples) window and 50% overlap between frames. The 256 points fast Fourier transform (FFT) is employed to estimate the magnitude spectrum of the windowed signal. The obtained noisy signal spectrum is divided into K sub-bands, and average value of the segmental signal-to-noise ratio (SEGSNR) is calculated over each preceding and succeeding k-th sub-band.

Moreover, spectral subtraction is implemented independently across multiple sub-band by subtracting the average value of the estimated magnitude of the noise spectrum in each k-th sub-band from the noisy signal spectrum using NES. It is observed that this approach prevents both over and under subtraction as well as signal distortion. Furthermore, the estimate of the enhanced output signal is obtained by the inverse fast Fourier transform (IFFT) of the enhanced spectrum using the phase of the original noisy spectrum. The resulting signal is overlap added to reconstitute the enhanced output signal.

4. Experimental Results and Analysis

The analysis carried out in this work is divided into two parts. In the first part, signals are simulated to observe performance of the proposed approach. Similarly, in the second part real-time signals are employed in the analysis.

4.1. Experiment 1: Simulation Based on Spectral Analysis

The three sinusoidal signals shown in Fig. (2a (blue), c (green), and e (cyan)) are simulated at 8 kHz sampling frequency. Their spectra are determined and are as depicted in Fig. (2b, d, and f). The spectrum of each simulated signal shows a mono-frequency spike as expected, occurring at the corresponding frequency of the simulated signal (i.e. 0.5 kHz (magenta), 1 kHz (black) and 1.5 kHz (red)). The result validates and shows that MATLAB® based tool is a viable means of obtaining precise results.

The three simulated sinusoidal signals are combined and a plot of the resulting composite signal (magenta) and the spectrum (red) are shown in Fig. 3. It shows three frequency spikes each located at the corresponding frequency.

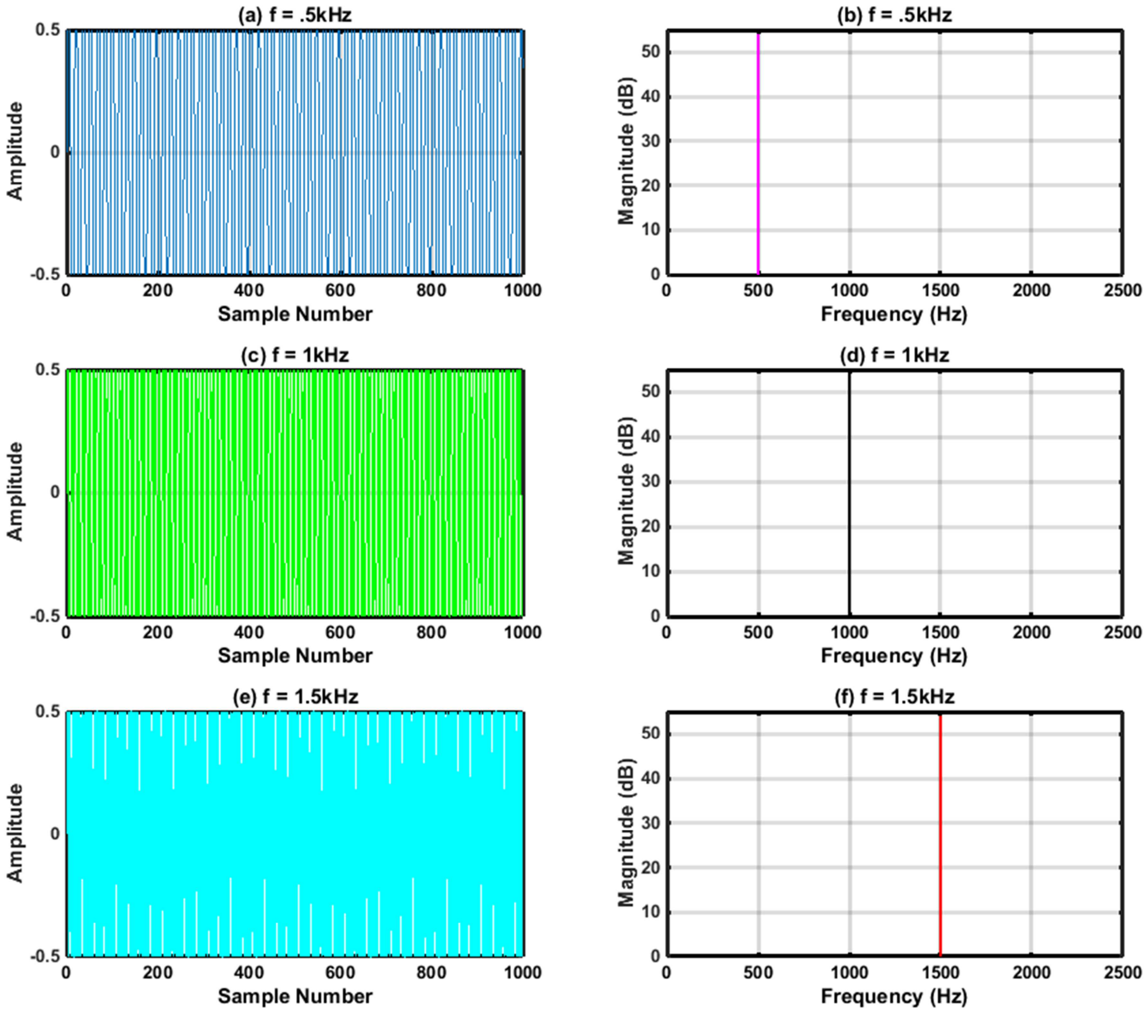

Fig. 2. Simulated Signals and the Spectra.

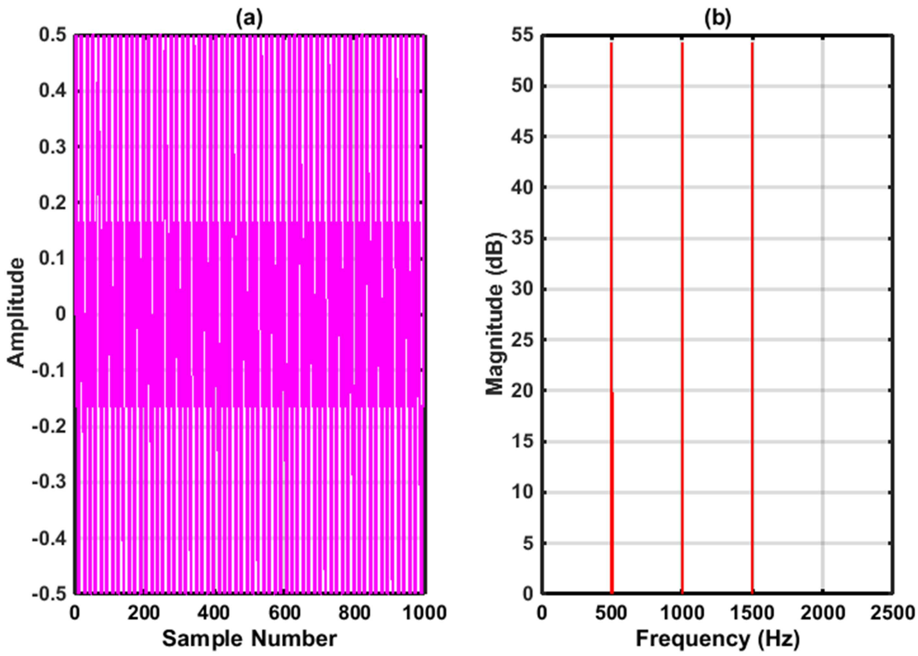

Fig. 3. Composite Signal and the Spectrum.

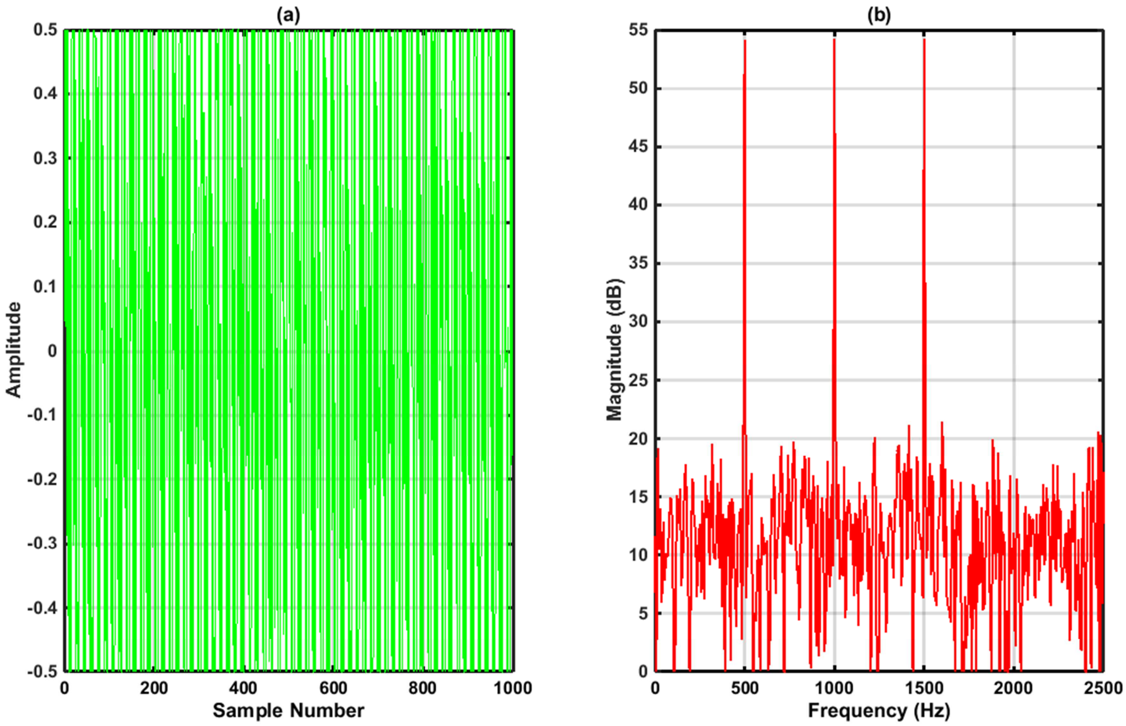

Fig. 4. (a) Composite Signal + Noise, (b) Spectra.

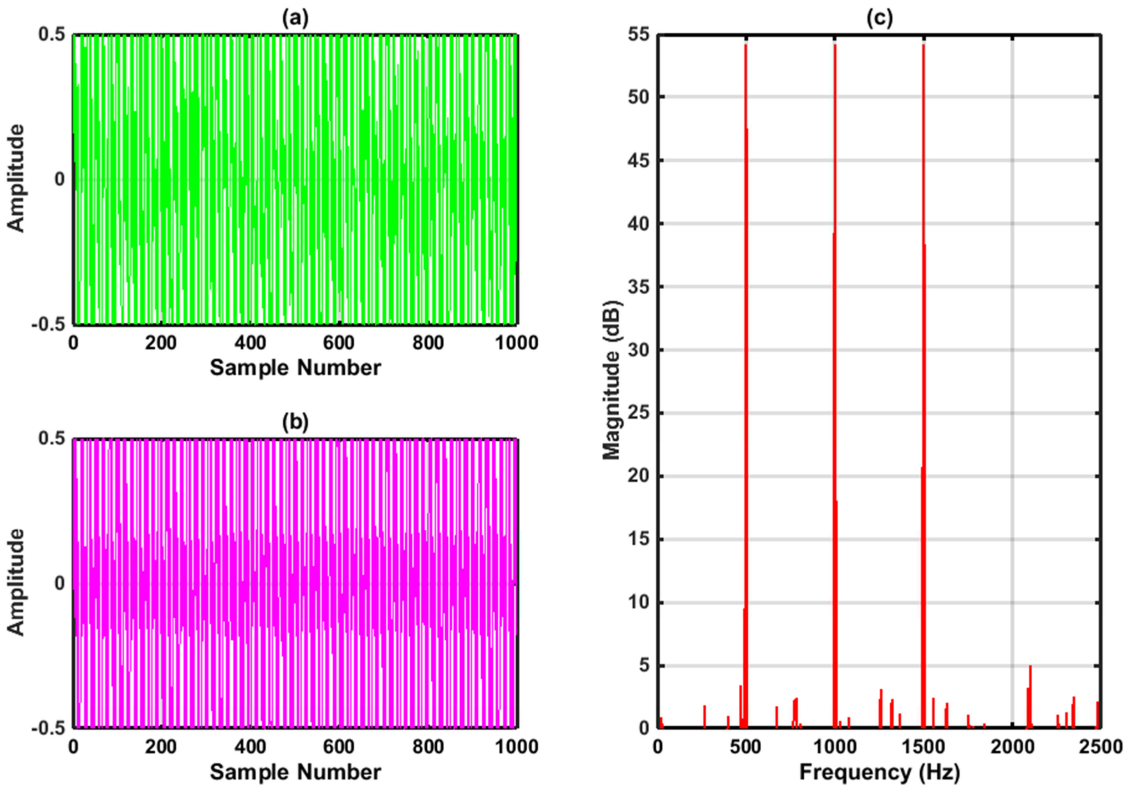

Fig. 5. (a) Composite Signal + Noise, (b) Restored signal and (c) Spectra.

Additionally, 1024 samples of simulated noise signal are added to the composite simulated signal to produce the noisy signal (green) and the spectrum (red) which are shown in Fig. (4a) and (4b) respectively. Furthermore, signal enhancement is implemented using conventional SS and the proposed multi-band SS technique at different SNRs. The corresponding plot of the noisy signal (green) and the enhanced signal (magenta) are shown in Fig. (5a) and (5b). The results of signal processing based on the conventional SS show some residual noise in the enhanced signal spectrum (red) which lead to signal distortion. This is depicted in Fig. (5c).

Furthermore, result shows that at the same SNR using the proposed multi-band SS technique, there are no notable anomalies as experienced in the conventional SS. Also, multi-band SS using proposed NES maintains high signal quality and consistently outperformed the conventional SS approach for all SEGSNRs considered. The plot of noisy signal (green), enhanced signal (magenta) and spectrum of restored signal (red) using MBSS are shown in Fig. (6a, 6b, and 6c) respectively.

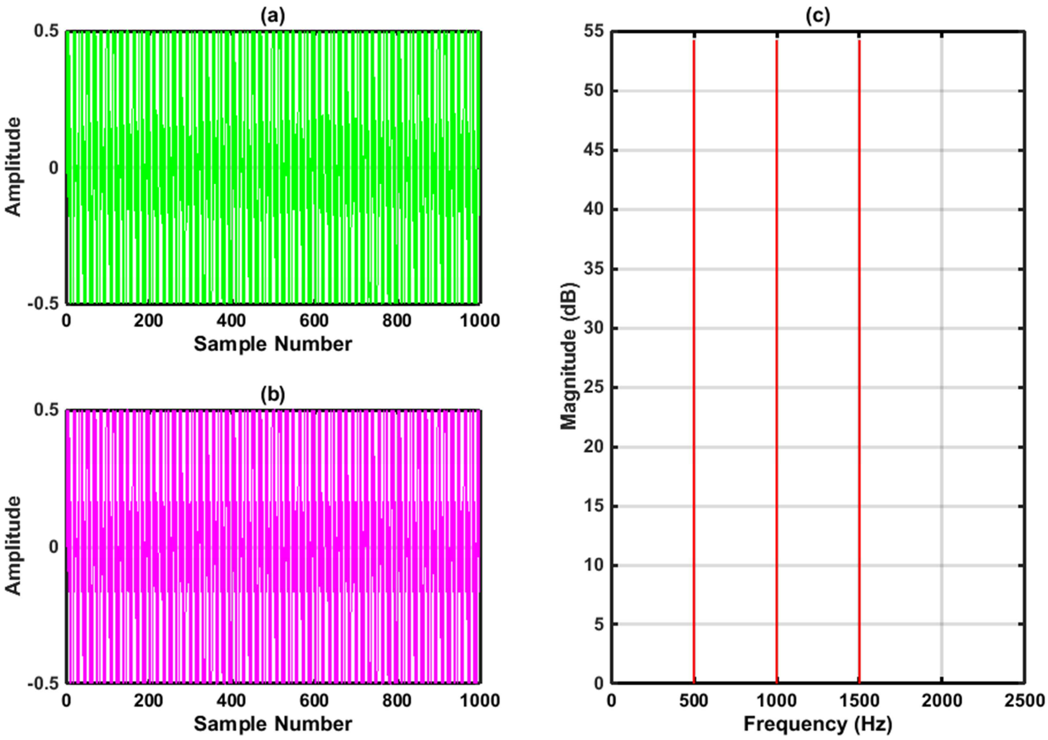

Fig. 6. (a) Composite Signal + Noise, (b) Restored signal and (c) Spectra.

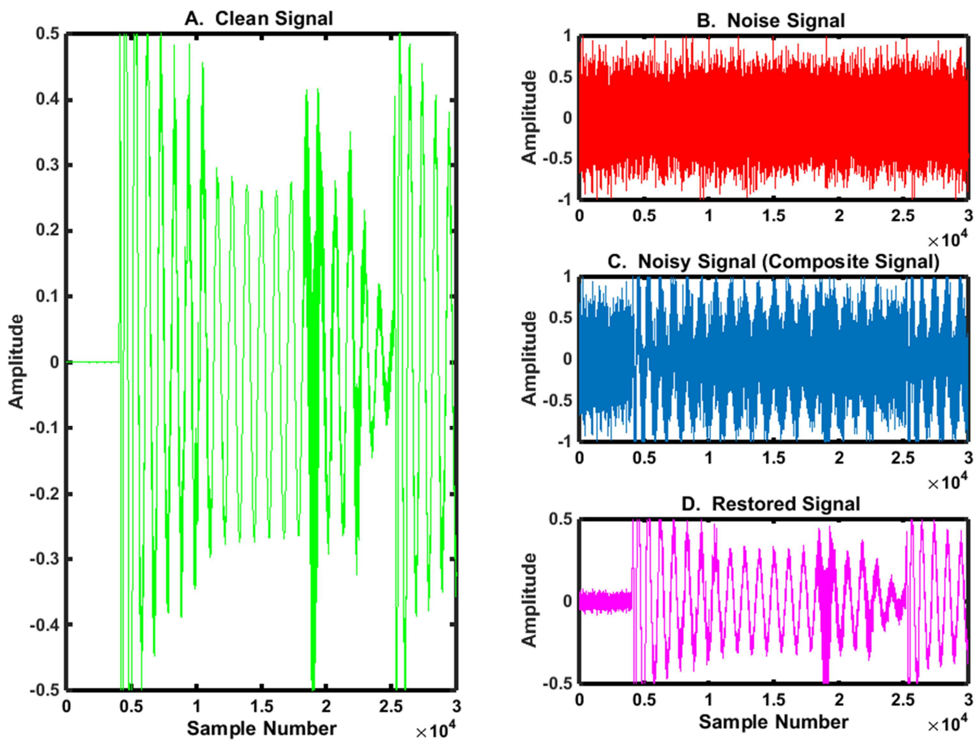

Fig. 7. (A) Clean Signal, (B) Noise Signal (C) Composite Noisy Signal and (D) Restored Signal.

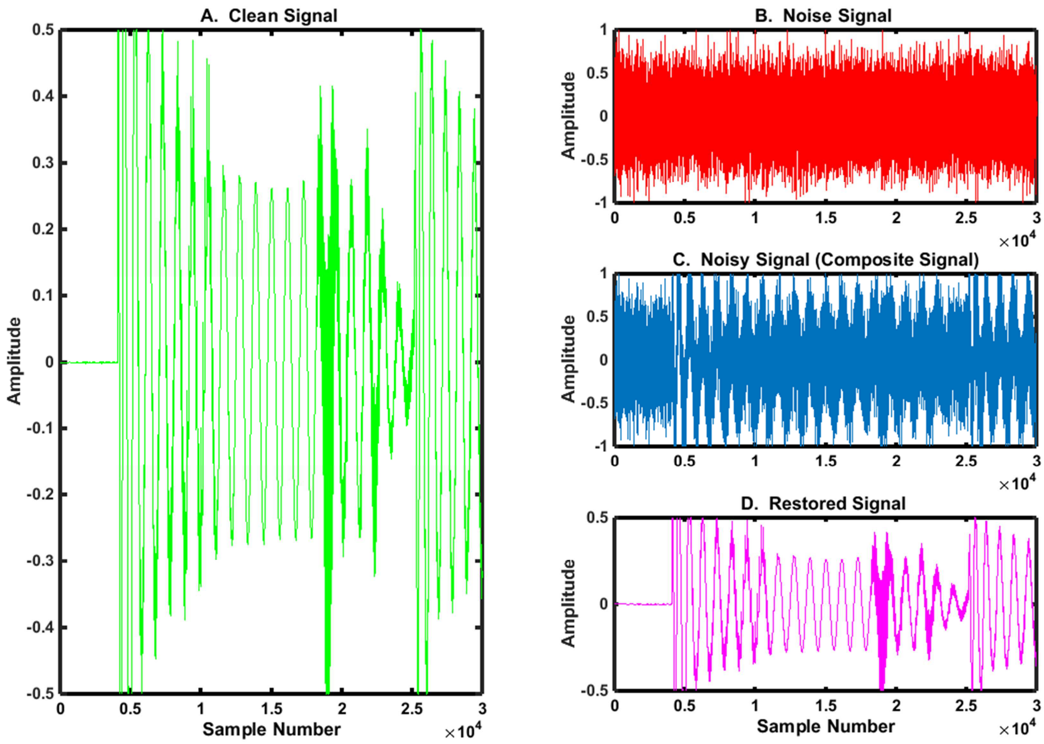

Fig. 8. (A) Clean Signal, (B) Noise Signal (C) Composite Noisy Signal and (D) Restored Signal.

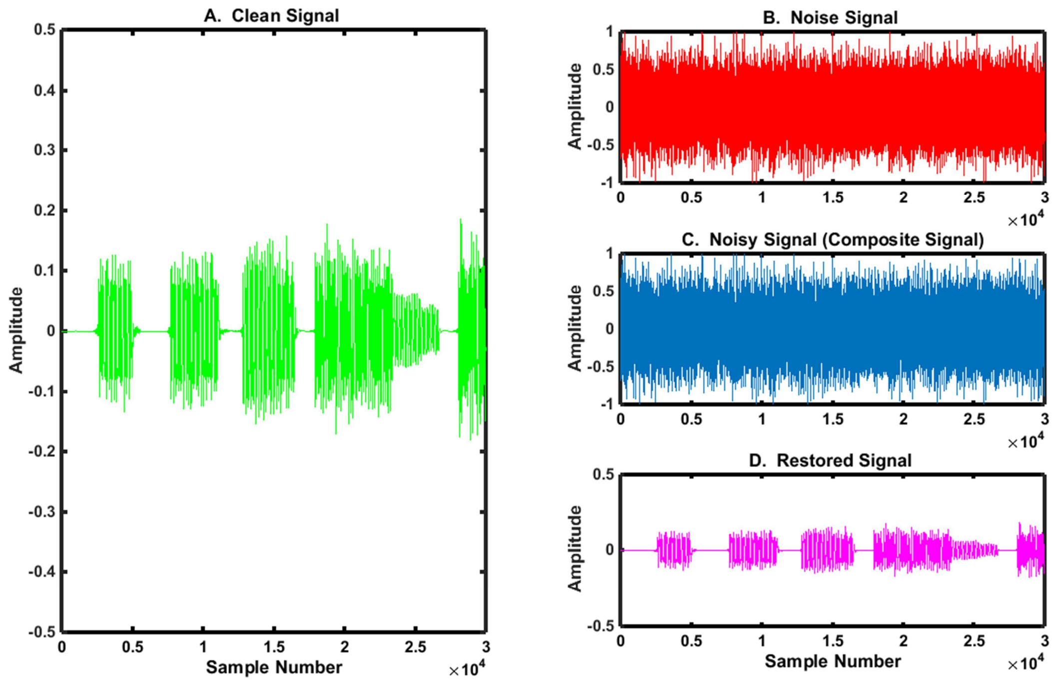

Fig. 9. (A) Clean Signal, (B) Noise Signal (C) Composite Noisy Signal and (D) Restored Signal.

4.2. Experiment 2: Simulation Based on Real-Time Signal Analysis

In this part, to confirm effectiveness of the results obtain in the first part, real-time signal and noise signal are employed. Clean signal and noise signal are combined to produce the noisy signal and signal enhancement analysis are implemented using conventional SS and the proposed multi-band SS technique. For the conventional SS approach, the corresponding plots of the clean signal (green), the noise signal (red), the noisy signal (blue) and the enhanced signal (magenta) are shown in Fig. (7A), (7B), (7C) and (7D). With reference to Fig. (7D), the result of signal processing based on the conventional SS shows some residual noise in the enhanced signal which give rise to signal distortion.

Furthermore, the same clean signal and noise signal are employed and the proposed multi-band SS technique which is based on NES is implemented. With this approach, result shows that there are no notable anomalies as experienced in the conventional SS. The plot of clean signal (green), the noise signal (red), the noisy signal (blue) and the enhanced signal (magenta) using MBSS are shown in Fig. (8A), (8B), (8C) and (8D) respectively.

Furthermore, three dissimilar clean speech signals with different spectral levels are added to samples of real-world noise. The obtained results are depicted in Fig. 9 - 11. Fig. A, B and C of each plot shows the clean signal (green), the noise signal (red) and the composite noisy signal (blue) respectively. The implementation of the proposed NES gives satisfactory enhanced speech, as shown in Fig. D of each plot that contains the restored signal (magenta).

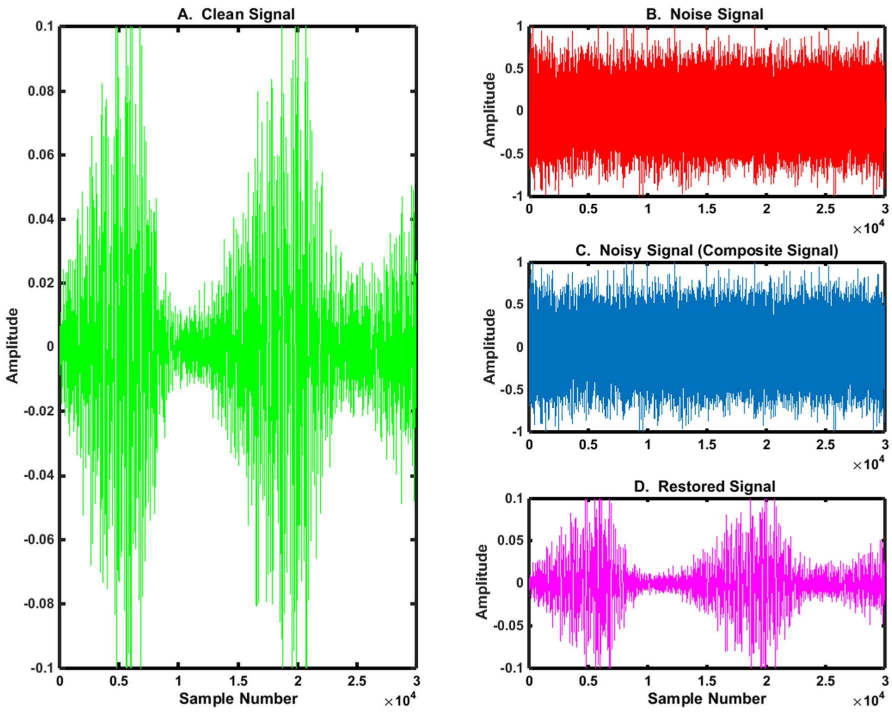

Fig. 10. (A) Clean Signal, (B) Noise Signal (C) Composite Noisy Signal and (D) Restored Signal.

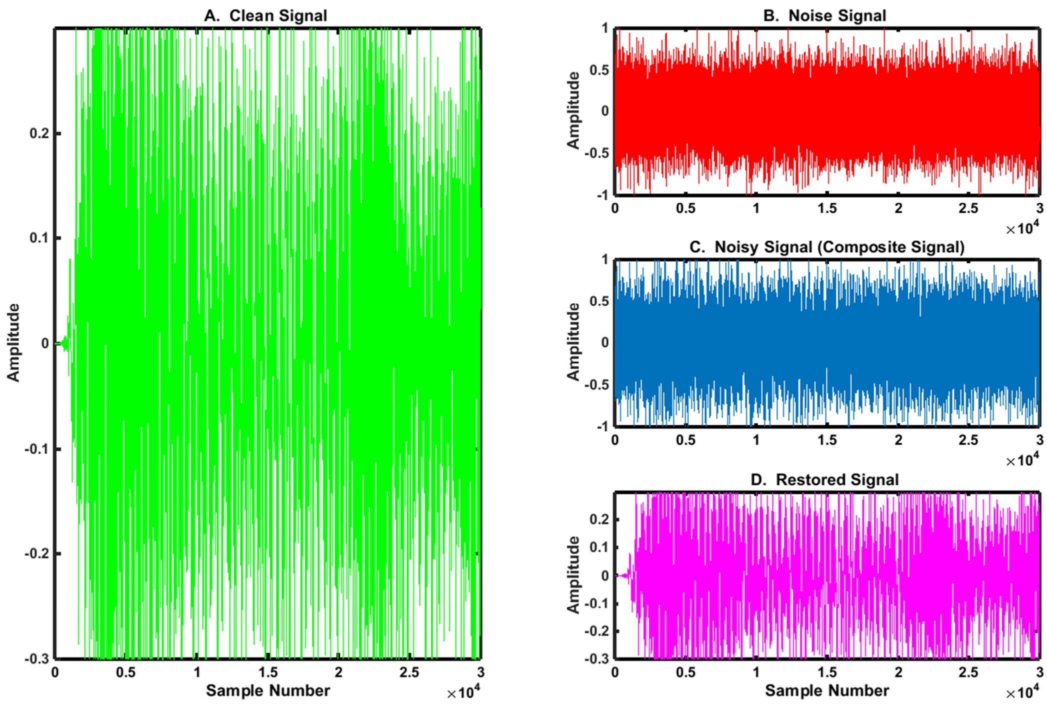

Fig. 11. (A) Clean Signal, (B) Noise Signal (C) Composite Noisy Signal and (D) Restored Signal.

The fact that digital speech processing specialists and radio broadcast professionals do not believe in a simple mathematical error criterion lead to further confirmation of the effectiveness of the proposed scheme by listening tests. Subjective listening tests are carried out to determine the quality and intelligibility of speech enhanced by the proposed scheme. The results of our test for residual noise and speech distortion show that the proposed NES based method outperforms the conventional approach.

5. Conclusion

In this paper, a novel multi-band spectral subtraction (MBSS) method is presented for enhancing speech signal corrupted by noise using the noise element suppression (NES) to prevent both over and under subtraction of noise signal in the composite signal. Also, it prevents signal distortion that is associated with the conventional method and offers comparatively higher signal-to-noise ratio. Therefore, the proposed approach relatively offers positive improvement with no adverse effect on the processed signal. The enhancement in the processed signal is due to the fact that the non-uniform effect of noise on the signal spectrum is taken into consideration over the entire speech spectrum.

References

[1] F. Jabloun and B. Champagne, “Incorporating the human hearing properties in the signal subspace approach for speech enhancement”, IEEE Trans. Speech Audio Process., vol. 11, no. 6, pp. 700–708. 2003.

[2] D. Burshtein and S. Gannot, “Speech enhancement using a mixture-maximum model”, IEEE Trans. Speech Audio Process., vol. 10, no. 6, pp. 341–351, 2002.

[3] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction”, IEEE Trans. Acoustics, vol. 27, no. 2, pp. 113–120, 1979.

[4] S. Doclo and M. Moonen, “GSVD-based optimal filtering for single and multi-microphone speech enhancement”, IEEE Trans. Sig. Process., vol. 50, no. 9, pp. 2230–2244, 2002.

[5] C. Zhixin, “Simulation of Spectral Subtraction Based Noise Reduction Method,” International Journal of Advanced Computer Science and Applications, vol. 2, no.8, pp. 30-32, 2011.

[6] G. Prabhakaran,; J. Indra and N. Kasthuri, “Tamil speech enhancement using non-linear spectral subtraction,” 2014 International Conference on Communications and Signal Processing pp.1482-1485, 2014.

[7] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, “Speech Enhancement based on Audible Noise Suppression”, IEEE Trans. Speech Audio Process., vol. 5, no. 6, pp. 497–514, 1997.

[8] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, “Musical noise generation analysis for noise reduction methods based on spectral subtraction and MMSE STSA estimatio,” in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’09), pp. 4433–4436, 2009.

[9] T. Inoue, H. Saruwatari, Yu Takahashi, K. Shikano, and K. Kondo “Theoretical Analysis of Musical Noise in Generalized Spectral Subtraction Based on Higher Order Statistics,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 6, pp. 1770-1779, 2011.

[10] I. A. Alimi and M. O. Kolawole, “Enhancement of Speech Communication Technology Performance a Using Adaptive-Control Factor Based Spectral Subtraction Method,” Journal of Telecommunication and Information Technology. vol.2, pp. 35-39, 2013.

Biography

Isiaka Ajewale Alimi Isiaka earns B.Tech. (Hons) and M.Eng. in Electrical and Electronics Engineering (Communication) from Ladoke Akintola University of Technology, Ogbomoso, Nigeria in 2001, and the Federal University of Technology, Akure, Nigeria in 2010 respectively. He is a Lecturer in the Department of Electrical and Electronics Engineering, Federal University of Technology, Akure, Nigeria. He has published 3 refereed international journals. He has extensive experience in radio transmission, as well as in Computer Networking. His areas of research are in Computer Networking and Security, Advanced Digital Signal Processing and Wireless communications. He is a COREN registered engineer.

Tusin Dayo Ebinowen received MBA and M.Eng. Degree in Communication Engineering from The Federal University of Technology, Ondo state, Nigeria in 2004 and 2012 respectively. He joined the services of the Nigerian National Petroleum Corporation as an Electrical Technician in the Technical Service Department from 1998-2000. He joined the Department of Electrical and Electronics Engineering, Federal University of Technology, Akure, Nigeria as Technologist in charge of Communication laboratory. His current research interests as a P.hD. student include, Radio frequency spectrum, Multi-cell cooperation in wireless networks. He is a COREN registered engineer.