Paper

Enhancement of Speech Communication Technology Performance Using Adaptive-Control Factor Based Spectral Subtraction Method

Isiaka A. Alimi[a,b] and Michael O. Kolawole[a]

[a] *Electrical and Electronics Engineering Department, Federal University of Technology, Akure, Ondo State, Nigeria*
[b] *Engineering Department, Positive FM, Federal Radio Corporation of Nigeria, Akure, Ondo State, Nigeria*

Abstract—This paper presents speech enhancement technique based on Spectral Subtraction (SS) method. SS is a renowned noise reduction technique that works on the principle that noise spectrum estimate over the entire speech spectrum can be subtracted from the noisy signal. On the contrary, most of the noise encountered in the real-world conditions is majorly colored. Unlike Additive White Gaussian Noise (AWGN), colored noise does not affect the speech signal uniformly over the entire spectrum. To mitigate effects of colored noise on the processed signal, we propose a Multi-Band Spectral Subtraction (MBSS) method using novel Adaptive-Control Factor (ACF). The spectrum is divided into frequency sub bands based on a nonlinear multi-band frame and various signal-to-noise ratios (SNRs) are considered. The proposed scheme results in better system performance with quality signal and unlike the basic SS method. It mitigates the effects of anomaly known as "musical" tones artifacts in the processed signal that result in residual noise and speech distortion. The computational complexity involved is minimal. Furthermore, simulation results show that the proposed algorithm removes more colored noise without removing the relatively low amplitude speech signal over the entire speech spectrum. Subjective listening tests, with clean speech signals and different noise levels, show discernable performance of our proposed method when compared with the conventional SS approach.

Keywords—Adaptive-Control Factor, MBSS, musical noise, subbands.

1. Introduction

Advances in digital signal processing have improved the quality of the existing and emerging communication technology services such as mobile telephony, teleconference systems, and Voice over Internet Protocol (VoIP). The corruption of speech signals due to presence of additive background and channel noise causes severe difficulties in various communication environments. Noise presence frequently degrades the quality of services and the information content of a signal [1]. To improve the quality of the corrupted signals, noise must be eliminated or suppressed. Noise suppression techniques are essential for these systems to operate efficiently [2].

In [3] Boll proposed Spectral Subtraction method of suppressing the effect of noise acoustically added to the speech signals. The approach is popular because of its simplicity and versatility in concept and effectiveness in enhancing speech degraded by additive noise [4]. The basic principle of the spectral subtraction method is to subtract the magnitude spectrum of noise from that of the noisy speech. The approach works under the assumption that noise signal is uncorrelated and additive to the speech signal [2]. While this power spectral subtraction method substantially reduces the noise levels in the noisy speech, it can cause deterioration of the recognition accuracy as well as introduce further distortion – called musical noise – in the speech signal [5], [6]. Musical noise consists of tonal remnant noise components that are annoyingly unpleasant to the ear.

Recent studies have focused on a nonlinear method to the subtraction process – justified by the variation of SNR across the enhanced speech spectrum [2], [7]. The spectrum of colored noise is not flat like the assumed white Gaussian noise. Consequently, the noise signal does not affect the speech signal uniformly over the whole spectrum. Certain frequencies are affected more adversely than others.

To prevent the variation of SNR across the enhanced speech spectrum and destructive subtraction of the speech while removing most of the residual noise, it is necessary to develop an appropriate factor that will subtract only the necessary amount of the noise spectrum from each frequency bin. In [8] criterion to quantify the amount of generated musical noise was proposed.

In this paper, a multi-band approach to spectral subtraction method that maintains a high speech quality and mitigates the stated anomalies using new Adaptive-Control Factor (ACF) is proposed. The ACF allows for the removal of less noise during relatively low amplitude speech and more noise during relatively high amplitude speech. The proposed approach divides the spectrum into frequency sub-bands based on a nonlinear multiband frame. For each sub-band, the noise corrupted speech spectrum in preceding and current time frames is compared to statistics of the noise spectrum to improve the determination of the speech activity in a given sub-band.

The mathematical descriptions of the MBSS and the proposed ACF are discussed in Section 2. Section 3 discusses the implementation of MBSS with ACF. Section 4 contains experimental results of the research. Conclusions are drawn in Section 5.

2. Multi-Band Spectral Subtraction

Suppose a clean signal $s(n)$ is corrupted by a stationary additive noise $d(n)$. The resulting received corrupted signal can be expressed as

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

where n is the discrete time index. The power spectrum of the received signal, at k instant, can be approximately estimated from:

$$|R(k)|^2 \approx |S(k)|^2 + |D(k)|^2. \tag{2}$$

The received signal is buffered and divided into segments of N samples length. Each segment is windowed, using Hamming window technique, and discretely Fourier transformed to N spectral samples. Windowing alleviates the effects of discontinuities at the endpoints of each segment and suppresses glitches. Therefore, it avoids the broadening of the frequency spectrum caused by the glitches [7], [9].

Following [3], the clean speech spectrum estimate is obtained as:

$$|\hat{S}(k)|^2 = |\hat{R}|^2 - \alpha|\hat{D}(k)|^2, \tag{3}$$

where α denotes an over-subtraction factor. This factor is for controlling the amount of noise subtracted from the noisy signal. For full noise subtraction, $\alpha = 1$ and for over-subtraction $\alpha > 1$.

A novel Adaptive-Control Factor $\alpha(k)$ is proposed that allows controlling mechanism within each frequency-band k, giving that noise is colored and has non-uniform spectral distribution. This ACF is scaled to accommodate for the multiple-frequency range that may exist in speech spectrum, expressed as:

$$\alpha(k) = \begin{cases} \left|\dfrac{f}{2\beta(k)}\right| & f \leq 2 \text{ kHz} \\ 1 & f > 2 \text{ kHz} \end{cases}, \tag{4}$$

where $\beta(k)$ is the normalized value of the noise spectrum dictated by the level of the signal. The $2\beta(k)$ accommodates for peak-to-peak consideration, and the frequency f is in kHz. The floor-noise may have approximate frequency as that of power-line interference and its harmonic component at about 50 Hz. The inclusion of frequency-components of $f < 50$ Hz is to accommodate the situation when the speech is contaminated by disturbances close to the signal being generated such as extragenoeous low-frequency, high-bandwidth components caused by body movement, and/or nearby processing equipment. Further, the border of $f \leq 2$ kHz reflects the limit where extraneous noise becomes problematic for normal speech recording range.

3. Implementation

The signal is first windowed using a 20 ms (160 samples) window and 50% overlap between frames. The magnitude spectrum of the windowed signal is estimated using 256 points Fast Fourier Transform (FFT) at 8 kHz sampling frequency. The noisy signal spectrum is divided into K sub-bands, and average value of the segmental SNR is calculated over each preceding and succeeding k-th sub-band. Then, spectral subtraction was implemented independently across multiple sub-bands by subtracting the estimated noise magnitude spectrum in each k-th sub-band from the noisy signal spectrum using ACF. This prevents both over and under subtraction as well as signal distortion. The estimated noise magnitude spectrum in each k-th sub-band is subtracted from the noisy signal spectrum. The processed k-th sub-bands are combined and then the enhanced estimate of the 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 output enhanced signal sequence. Different noise scenarios were considered with variable intensity and sub-band variable frequencies to test the effectiveness of MBSS technique.

4. Experimental Results

4.1. Simulation Results

Firstly, a real-world low-level noise scenario environment like home or office is considered. In this situation, $4.5 \cdot 10^4$ samples of real-world noise are added to the same value of clean speech signal, as shown in Figs. 1a-c the composite noisy signal. The implementation of the proposed MBSS gives satisfactory enhanced speech, as seen in Fig. 1d.

Furthermore, a real-world medium level noise scenario like campus environment is considered. In this condition, $4.5 \cdot 10^4$ samples of medium level noise are added to the same value of clean speech signal, as shown in Figs. 2a-c the composite noisy signal. Fig. 2d depicts enhanced speech obtained with the implementation of MBSS.

Additionally, this paper further examined a high-level noise environment to experiment effectiveness of the proposed approach. A real-world high-level noise scenario like manufacturing company is analyzed. In this environment, noise emanates from different sources like heavy duty generator and production machines. In this situation, $4.5 \cdot 10^4$ samples of real-world noise are added to the same value of clean speech signal, as shown in Figs. 3a-c the composite noisy signal. The implementation of MBSS gives satisfactory enhanced speech, as seen in Fig. 3d. In addition,

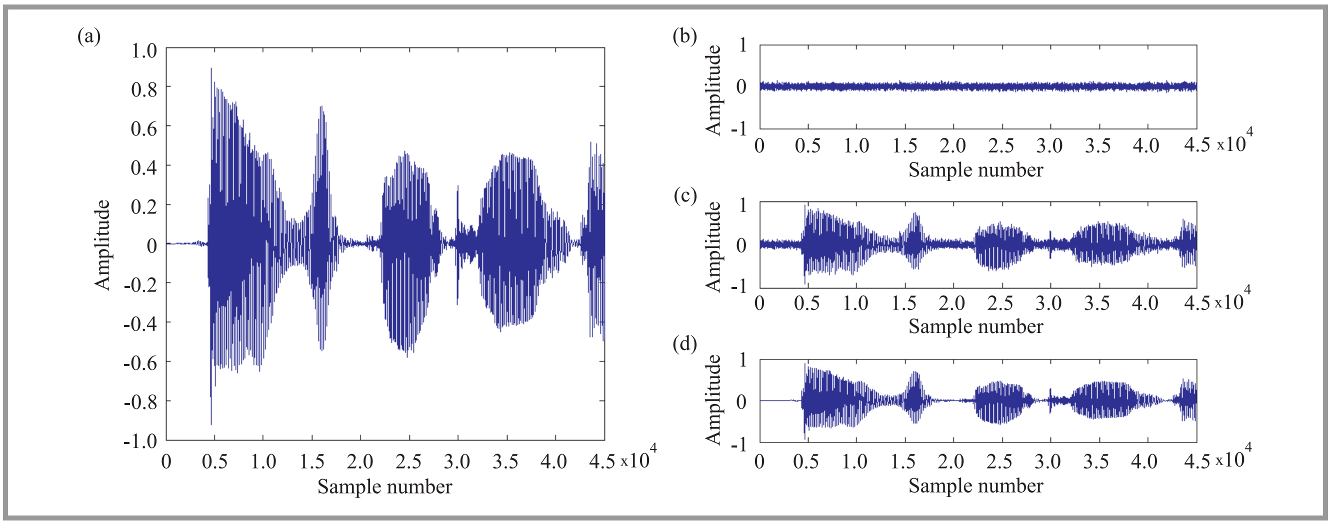

Fig. 1. Plots of (a) clean signal, (b) low level noise signal, (c) noisy signal and (d) restored signal.

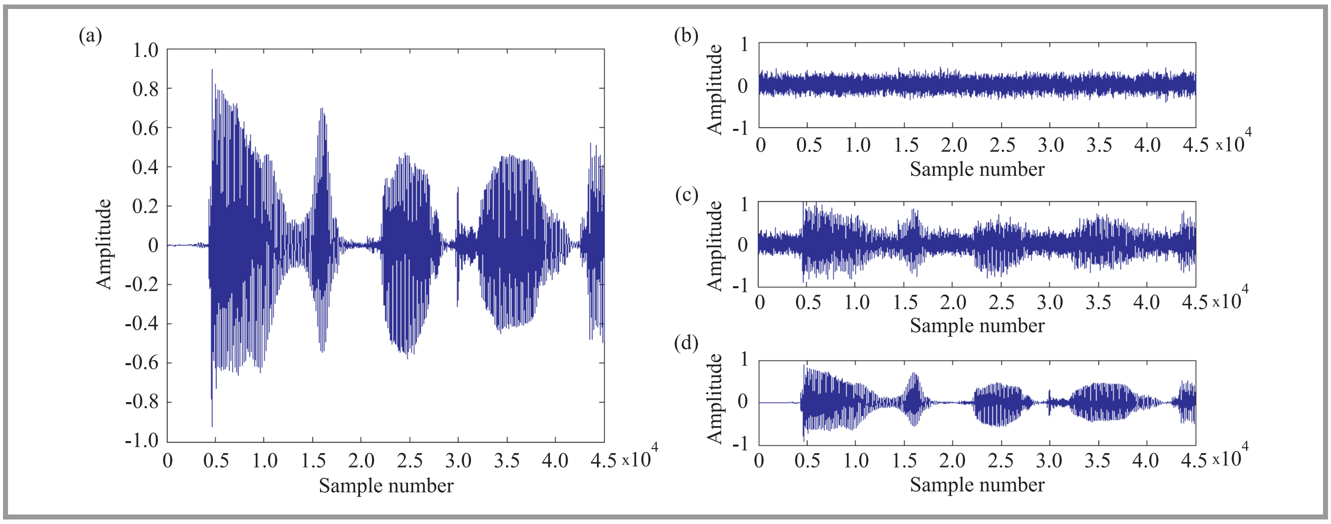

Fig. 2. Plots of (a) clean signal, (b) low level noise signal, (c) noisy signal and (d) restored signal.

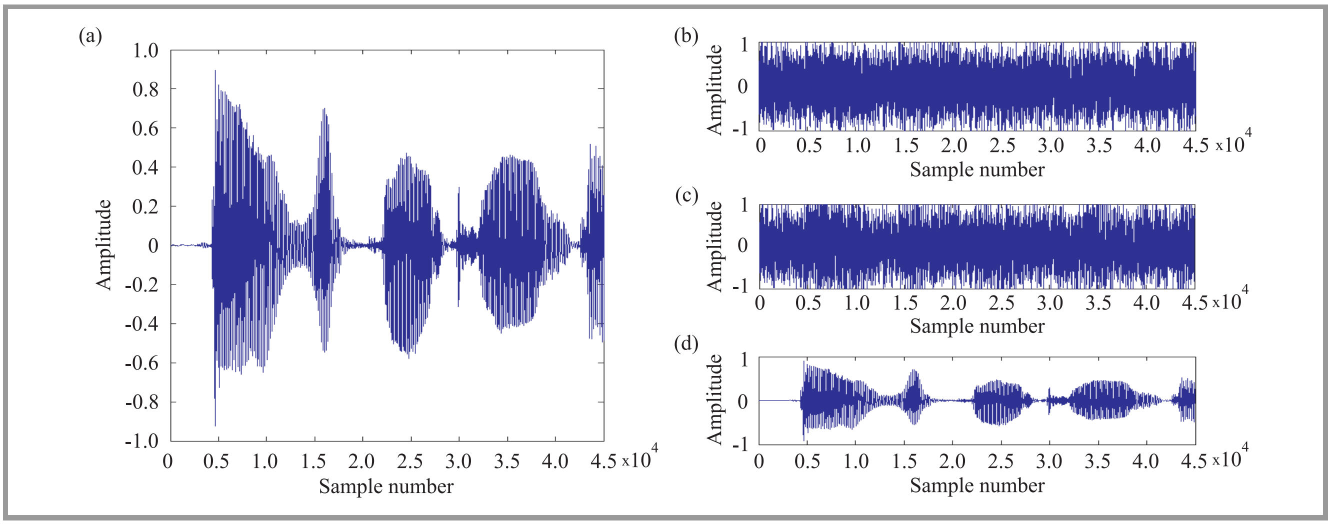

Fig. 3. Plots of (a) clean signal, (b) low level noise signal, (c) noisy signal and (d) restored signal.

the proposed algorithm removes more colored noise without removing the relatively low amplitude speech signal over the entire speech spectrum.

4.2. Listening Test Results

The human listener does not believe in a simple mathematical error criterion. As such, in order to confirm the effectiveness of results obtained from simulations for the proposed method, subjective listening experiments were carried out with clean speech signals and different noise levels. The sampling frequency for all recordings was 8 kHz. 12 persons took part in the listening tests carried out to determine subjective quality and intelligibility of speech enhanced by our method. Eight of the participants are radio broadcast professionals who has about 8 years experience in both analogue and digital speech processing and are in their early thirties. Furthermore, four students working on digital speech processing area and in their twenties participated in the test.

Participants were told to choose the signal they preferred from the ACF-based and conventional SS approaches, as well as choosing according to how intelligible and quality the signal is. The results of our test for residual noise for real-world low-level noise shows that 6 persons preferred ACF approach, 3 persons preferred conventional SS approach, while 3 persons are indifferent. In addition, for residual noise for real-world medium level noise, results show that 8 persons preferred ACF approach, 3 persons preferred conventional SS approach, while 1 person is indifferent. Furthermore, test for residual noise for real-world high-level noise shows that 10 persons preferred ACF approach and 2 persons preferred conventional SS approach. Table 1 shows percentage representation of the residual noise result obtained.

Table 1
The test results for residual noise

Noise type	ACF based MBSS [%]	Conventional SS [%]	Indifferent [%]
Low level	50	25	25
Medium level	67	25	8
High level	83	17	0

Table 2
The test results for speech distortion

Noise type	ACF based MBSS [%]	Conventional SS [%]	Indifferent [%]
Low level	67	25	8
Medium level	83	17	0
High level	92	8	0

The results of test for speech distortion for real-world low-level noise show that 8 persons preferred ACF approach, 3 persons preferred conventional SS approach while 1 person was indifferent. In addition, results of test for speech distortion for real-world medium level noise shows that 10 persons preferred ACF approach, 2 persons preferred conventional SS approach. Furthermore, test for speech distortion for real-world high-level noise shows that 11 persons preferred ACF approach, 1 person preferred conventional SS approach. Table 2 shows percentage representation of the speech distortion result obtained. These results show that the proposed ACF based method outperforms the conventional SS approach.

5. Conclusion

This paper has presented a novel Multi-Band Spectral Subtraction method for enhancing signal corrupted by noise. The introduction of ACF prevents both over and under subtraction as well as signal distortion. In addition, listening test results show that the proposed method performs better than the conventional SS approach. Our approach maintains high signal quality and offers positive improvement that consistently outperforms the conventional spectral subtraction approach for all SNRs observed with no adverse effect on the processed signal. The improvement is because the non-uniform effect of colored noise on the signal spectrum is taken into consideration. This results in a comparatively higher SNR.

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, 2003, pp. 700–708.

[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 multimicrophone speech enhancement", *IEEE Trans. Sig. Process.*, vol. 50, no. 9, pp. 2230–2244, 2002.

[5] Y. Hu and P. C. Loizou, "A perceptually motivated approach for speech enhancement", *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 5, pp. 457–465, 2003.

[6] J. A. Haigh and J. S. Mason, "Robust voice activity detection using cepstral features", *IEEE Tencon*, vol. 3, pp. 321–324, 1993.

[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, "Automatic Optimization Scheme of Spectral Subtraction Based on Musical Noise Assessment Via Higher-Order Statistics", in *Proc. 11th Int. Worksh. Acoustic Echo and Noise Control IWAENC 2008*, Seattle, Washington, USA, 2008, pp. 1–4.

[9] R. L. Fante, *Signal Analysis and Estimation*. Toronto: Wiley, 1988.

Isiaka Ajewale Alimi received B. Tech. (Hons) and M. Eng. in Electrical and Electronics Engineering respectively from Ladoke Akintola University of Technology, Ogbomoso, Nigeria in 2001, and the Federal University of Technology, Akure, Nigeria in 2010. He is currently pursuing his Ph.D. at the Federal University of Technology Akure. 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 (Council for the Regulation of Engineering in Nigeria) registered engineer, a member of the Nigerian Society of Engineers (NSE).
E-mail: compeasywalus2@yahoo.com
Electrical and Electronics Engineering Department
The Federal University of Technology
P.M.B. 704, Akure, Ondo State, Nigeria

Michael O. Kolawole received B. Eng. (Victoria University, Melbourne 1986) and Ph.D. (UNSW, 2000) in Electrical Engineering, and Master of Environmental Studies (Adelaide, 1989). He is concurrently LEAD Scholar and Professor of Electrical Engineering (Communication) at the Federal University of Technology, Akure Nigeria and Director of Jolade Consulting Company (Melbourne Australia) where, since its establishment, he has provided vision and leadership. He has published over 40 peer-reviewed papers, holds 2 patents and has overseen a number of operational innovations. Mr. Kolawole is the author of three books and co-author of fourth. He has consulted widely and published extensively in his areas of expertise. His research interests are in biomedical engineering, satellite communication engineering, radar systems and tracking, and remote sensing.
Electrical and Electronics Engineering Department
The Federal University of Technology
P.M.B. 704, Akure, Ondo State, Nigeria