Signal Processing: Musical Noise Reduction By Processing Spectrogram of Spectral Subtraction Output

ICSV 15
15th International Congress on Sound and Vibration
6-10 July 2008, Daejeon, Korea

MUSICAL NOISE REDUCTION BY PROCESSING SPECTROGRAM OF SPECTRAL SUBTRACTION OUTPUT

Hamid Reza Abutalebi[1] and Behdad Dashtbozorg[2]

[1,2] Signal Processing Research Lab, Electrical Engineering Department, Yazd University
Yazd, Iran
habutalebi@yazduni.ac.ir

Abstract

This paper addresses a novel method for musical noise reduction. Musical noise is undesired result of using spectral subtraction algorithm for de-noising speech signals. The existence of musical noise makes the speech signal unnatural and annoying for human hearing system. The proposed method applies image-processing techniques on the spectrogram of spectral subtraction output. To detect musical noise blocks, we present a method that is based on the symmetry in the adjacency of each musical noise block. To assess the performance, we have used two objective measures: Segmental SNR (SegSNR), and Log-Likelihood Ratio (LLR) distance. We have also examined the comparative results using informal subjective listening tests. It has been shown that the proposed method effectively suppresses musical noise without noticeable effect on speech intelligibility.

1. INTRODUCTION

The main objective of speech enhancement is to reduce the corrupting noise component of noisy speech signal while preserving the original speech quality as much as possible. Spectral subtraction is one of the earliest methods for noise reduction. This method enhances the speech signal by subtracting the noise spectrum estimation from noisy signal spectrum [1]. Since the noise estimation is not accurate, in some frames, the estimation of (clean) speech spectrum becomes negative. This negative value is actually replaced by zero (or a small positive value); this, in turn, raises another undesired noise in speech signal that makes it unnatural [2]. This noise is called *musical noise*. The intensity of musical noise is depending on the parameters of the spectral subtraction method. In the literature, several solutions have been proposed for the musical noise problem [3-5]. Most of these techniques do not have the desired efficiency in low Signal to Noise Ratio (SNR). Some of them reduce musical noise just a little, and the others destroy the desired signal, which in turn reduces the speech intelligibility.

In this research, we have proposed a method that applies image processing techniques on the spectrogram of spectral subtraction output. We scan and analyze this spectrogram in 3*3 blocks. To detect musical noise blocks, we present a method based on the symmetry in the adjacency of each musical noise block. The time-frequency blocks around each musical noise block are symmetric in quantity and this characteristic is used in our algorithm to detect musical noise blocks.

In section 2, we explain the spectral subtraction method and the musical noise. The characteristics of musical noise time-frequency blocks and the symmetry in these blocks are described in section 3. In section 4, we compare some techniques for musical noise reduction. The proposed method is explained in section 5. In section 6, we compare the performance of this method with two previously-reported techniques via two objective measures: 1) Segmental SNR (SegSNR), and 2) Log-Likelihood Ratio (LLR) distance. Finally, we have concluding remarks in section 7.

2. SPECTRAL SUBTRACTION METHOD AND MUSICAL NOISE

Clean speech corrupted by additive noisy signal can be expressed as follows:

$$x[n] = s[n] + n[n] \tag{1}$$

where $x[n]$ is the noisy speech signal, $s[n]$ is the clean speech signal and $n[n]$ is the noise signal. In the Short-Time Fourier Transform (STFT) domain, we have:

$$X(kl,\omega) = S(kl,\omega) + N(kl,\omega) \tag{2}$$

$X(kl,\omega)$, $S(kl,\omega)$ and $N(kl,\omega)$ are STFT of noisy speech, clean speech and noise signals, respectively; in these STFTs, ω is frequency bin index, k is time frame index and l is sampling frequency. For estimating the clean signal, we should estimate the noise spectrum. This estimation is obtained during the silences that are determined by means of a Voice Active Detector (VAD). Finally, the spectrum of clean speech is estimated through:

$$|S(\omega)| = \left(|X(kl,\omega)|^{\alpha} - \beta \cdot E\left\{|N(kl,\omega)|^{\alpha}\right\}\right)^{\frac{1}{\alpha}} \tag{3}$$

where α and β are the parameters of spectral subtraction method. As the last step, by adding the phase of noisy signal and inverse STFT, we estimate the (time-domain) clean signal as follows:

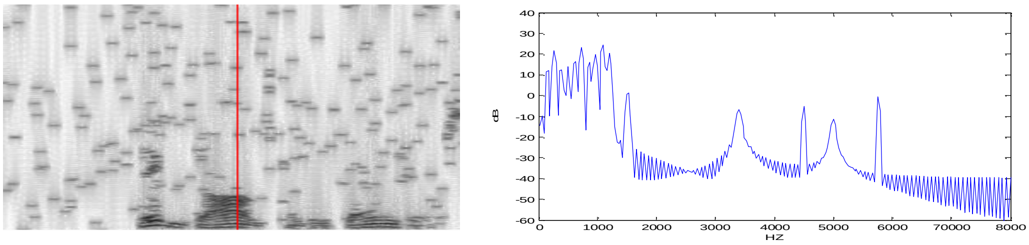
$$\hat{s}(n) = F^{-1}\left\{|\hat{s}(kl,\omega)| \exp(j\angle X(kl,\omega))\right\} \tag{4}$$

Incorrect estimation of noise spectrum makes the result of equation (3) negative. This (incorrect) value is actually replaced by zero; the change of some frequency component into zero produces a residual noise that is called musical noise. Musical noise mostly occurs in low-energy frames (i.e., during silences and unvoiced phonemes). To decrease this residual noise, we should firstly detect the time-frequency blocks that include musical noise and then, change them somehow to reduce the unnatural voice.

The spectral subtraction method (equation (3)) has two parameters α and β that control the method [6]. α affects on speech intelligibility, whereas β controls the amount of noise suppression. By using large α, we can increase the intelligibility. On the other side, large β attenuates the unwanted noise. However, this also leads to weakening of speech contents (especially in unvoiced phonemes) and consequently, reducing the intelligibility [6]. There are several methods for obtaining the optimal values of α and β. In this research, we evaluate the musical noise reduction methods with various values of α and β, in order to find the optimal results.

3. SPECTROGRAM AND CHARACTERISTIC OF MUSICAL NOISE BLOCKS

To draw the spectrogram, we firstly divide the speech signal into frames and then, apply the Fourier transform on each frame. Next, the amplitude of these STFTs are scaled and interpreted in grey level; this constitutes a column of the spectrogram. The STFTs of successive frames placed together and construct the whole spectrogram. The spectrogram of spectral subtraction output have been drawn in figure 1(a), where the musical noise is noticeable as discrete pixel blocks.



(a) (b)

Figure 1. (a) Spectrogram of speech signal with musical noise, (b) spectral curve of one frame

As discussed above, incorrect noise estimates result in zero frequency components. In turn, the presence of these zero blocks and a few non-zero blocks in one frame make the output signal to act as a cosine function (with variable frequency). In figure 1(b), we have drawn the STFT of a typical frame. In this figure, delta-like-peaks display the effect of the musical noise.

4. MUSICAL NOISE REDUCTION BASED ON SIGNAL SPECTROGRAM

Several methods have been already proposed for musical noise reduction, some of them work on output spectrogram. Whipple [3] has presented a method that acts based on the blocks energy level [3]. In this technique, a 3*3 block and a 7*7 block have been considered around each uncertain musical noise block. If the power of 3*3 block is noticeably more than 7*7 one, the centroid block will be considered as musical noise and changed toward zero. This method fails in the detection of musical noise blocks when they occurred in adjacent blocks.

In [4] another method was proposed for musical noise reduction. In this method for every uncertain musical noise block, some concentric blades were considered. Then, by comparing the variances of the blades with a pre-defined threshold, the presence of musical noise would be confirmed or denied. Similar to the previous one, this method has not enough proficiency in sever musical noise presence; in addition, this method is too complicated and time consuming.

5. PROPOSED METHOD FOR MUSICAL NOISE REDUCTION

In this section, we propose a novel method for musical noise reduction. In this method, we use the visual characteristics of signal spectrogram as an image. Through a detailed version of the spectrogram of the figure 1(a), we can see the musical noise as some 3*3 blocks. To clarify the proposed method, we simply consider a typical (uncertain) musical noise block in figure 2.

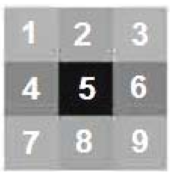


Figure 2. One 3*3 musical noise block

By comparing the adjacent pixels of musical noise pixel (pixel 5), it is obvious that the upper pixel (pixel2) has the same value as the lower pixel (pixel8). This symmetry around the main frequency of musical noise is due to the presence of delta-like-peaks in signal spectrum (as discussed in section 3). We use this fact for the detection of musical noise blocks as follows. If for any pixel q, we have:

$$q(i+1, j) = q(i-1, j) \tag{5}$$

then, the pixel is considered as musical noise and its value is replaced by zero (or a pre-defined small value).
In some cases, musical noise occurs in neighboring blocks (see figure 3). In such cases, the adjacent frequency components of these blocks (upper and lower pixels) are overlapped, so their value will change and there will be no more symmetry.

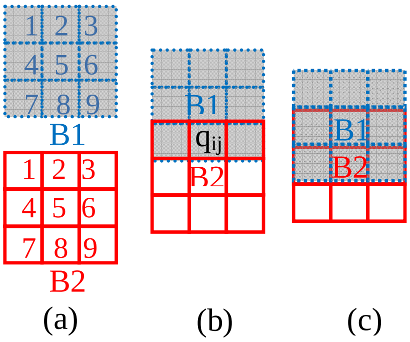


Figure 3. (a) Separate blocks, (b) one row overlapped blocks, (c) two row overlapped blocks

In the case of figure 3(b), the musical noise blocks are overlapped only in one row; if we consider one of these 3*3 blocks as $\{B_1(n)\ , n=1,2,\ldots 9\}$ and the other one as $\{B_2(n)\ , n=1,2,\ldots 9\}$, according to the symmetry around main frequency of musical noise, for each of these blocks we have:

$$B_i\ (2) = B_i\ (8) \quad , \quad i = 1,2 \tag{6}$$

Due to overlapping, some pixels of these two blocks (e.g., $B_2(2), B_1(8)$) will present identical actual frequency components; thus the observed pixel $q(i,j)$ has a value which is equal with the sum of the primary values of $B_2(2)$ and $B_1(8)$:

$$q(i, j) = B_1\ (8) + B_2\ (2) \tag{7}$$

According to equation (6) and (7), we have:

$$q(i, j) = q(i-2, j) + q(i+2, j) \tag{8}$$

For each pixel, if the equation (8) is satisfied, pixel will be considered as musical noise and its value will change to zero (or a small pre-defined value).

In the case of figure 3(c), where the musical noise blocks have been overlapped in two rows, there is no mathematical solution. To detect this kind of musical noise blocks, we propose another solution based on the energy of the surrounding 3*3 blocks. According to the figure 4, we consider four 3*3 blocks in the adjacency of each (uncertain) 3*3 musical noise block.

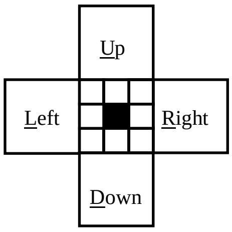


Figure 4. The 3*3 adjacent blocks of musical noise block

For each 3*3 block (see figure 4), average energy of the pixels is computed; the block is considered as a musical noise block if we have:

$$E(B_C) < \gamma.[E(B_U) + E(B_D) + E(B_L) + E(B_R)] \quad (9)$$

where *E[.]* is averaging function and γ is a constant value.

6. SIMULATIONS

For simulation, we use a clean speech signal sample (with 8 second length and 16 kHz sampling frequency). This signal becomes noisy with white Gaussian noise at various SNRs (-10dB, -5dB, 0dB, 5dB, 10dB, 15dB, 20dB). After that, we apply spectral subtraction method on these sample noisy signals with $\alpha=2$ and various β ($\beta=1, 2, 4$).

We compare the results via two objective measures: SegSNR and LLR distance. SegSNR is computed by calculating the SNR of each frame and averaging over all frames. To compute the LLR distance, we firstly calculate the LPC coefficients [7]. For each frame that the LPC vector of clean signal is $\vec{a}_\phi$, the autocorrelation matrix of clean signal is R_ϕ and LPC vector of processed signal is $\vec{a}_d$, LLR distance is determined as follows:

$$d_{LLR}(\vec{a}_\phi, \vec{a}_d) = \log\left(\frac{\vec{a}_\phi R_\phi \vec{a}_d^T}{\vec{a}_d R_\phi \vec{a}_\phi}\right) \quad (10)$$

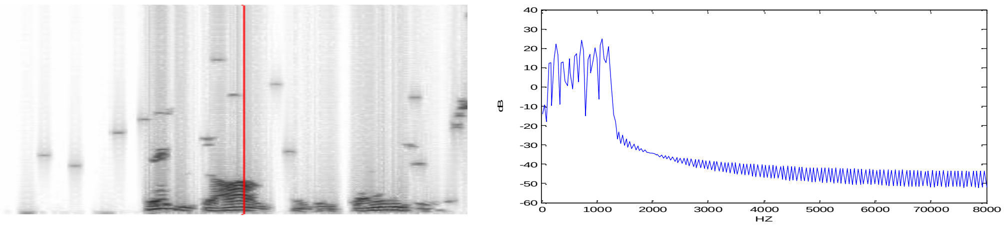
The total LLR value is computed from averaging over all frames; it is noticeable that LLR value has inverse relation with signal quality.

In section 6.1 we compare the performance of our proposed method in various values of β. In section 6.2, this method is compared with state-of-the-art methods.

6.1. Performance Comparison in Various Values of β

The figure 5 shows the result if applying our method on signal of figure 1. (figure 1 was the output of spectral subtraction method with InputSNR = -5 dB, $\alpha=2$, $\beta=4$). It is obvious that the delta-like-peaks of frequency curve have been disappeared and musical noise has been effectively removed.

In the next step, to find the optimal value of β, we apply the proposed method on outputs of spectral subtraction with $\alpha=2$ and $\beta=1, 2, 4$. The test is repeated for various Input SNRs. The results of this simulation are shown in table 1 and figures 6, 7.



(a) (b)

Figure 5. (a) Spectrogram after applying image processing method, (b) spectral curve of one frame

Table 1. The comparison of SegSNR and LLR distance measures after image processing method with various value of β

Overall Noisy Input SNR	Segmental SNR				LLR			
	Noisy Signal	β=1	β=2	β=4	Noisy signal	β=1	β=2	β=4
20dB	9.00	12.49	11.27	9.33	1.96	1.31	1.32	1.29
15 dB	5.54	10.61	9.67	8.82	2.23	1.32	1.30	1.30
10 dB	2.24	5.83	8.38	7.55	2.52	1.39	1.31	1.33
5 dB	-0.79	2.16	3.64	6.33	2.85	2.40	1.33	1.27
0 dB	-3.52	-1.27	0.40	1.93	3.18	2.73	2.72	1.34
-5 dB	-5.42	-3.96	-2.32	-0.46	3.49	3.23	3.08	1.87
-10 dB	-6.81	-6.06	-3.48	-2.78	3.77	3.60	3.45	2.47

According to table 1 and figures 6 and 7; the superiority of the proposed method is more obvious in “low SNRs and high β” or “high SNRs and low β”. Considering more attention on low SNRs, we choose $\alpha=2$ and $\beta=4$ for next experiments.

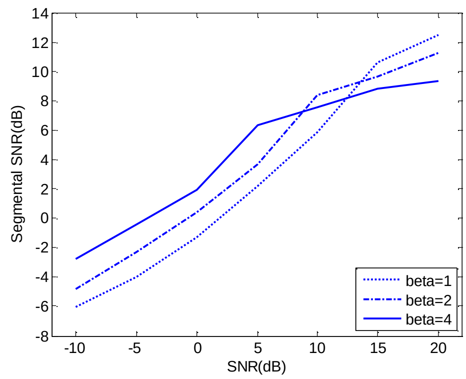


Figure 6. The comparison of SegSNR for the outputs of proposed method (applied on spectral subtraction method with various value of β)

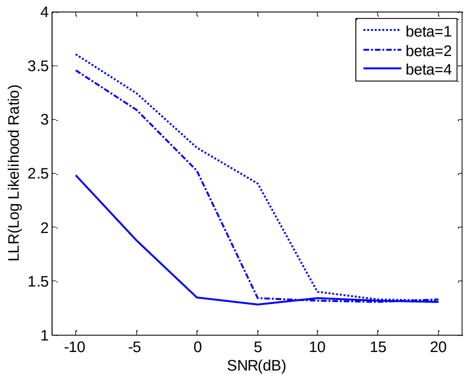


Figure 7. The comparison of LLR distances for the outputs of proposed method (applied on spectral subtraction method with various value of β)

6.2. The Comparison of Proposed Method with Other State-of-the-art Methods

In this experiment, the proposed (image processing) method is compared with two spectrogram-based methods for musical noise reduction: 1) energy threshold, and 2) blade variance (as explained in section 4). We have also added the results for noisy signal and the spectral subtraction method (without post-processing) for comparison. Spectral subtraction parameters has been set to $\alpha=2$, $\beta=4$. The results have been shown in tables 2 and 3.

Table 2. The SegSNR comparison of proposed method with energy threshold method, blade variance method, spectral subtraction method and noisy signal

Overall Input SNR	Segmental SNR				
	Noisy Signal	**Spectral Subtraction**	**Energy Threshold**	**Blade Variance**	**Image Processing**
20dB	9.0006	9.3316	9.0666	11.3209	8.1346
15 dB	5.5455	8.8249	8.4087	8.2971	7.8422
10 dB	2.2461	7.5505	7.1821	5.5451	7.1811
5 dB	-0.7949	6.3311	6.0188	3.2958	6.6423
0 dB	-3.5215	1.9310	2.0979	1.6916	5.7852
-5 dB	-5.4204	-0.4679	-0.2570	1.0915	2.3350
-10 dB	-6.8159	-2.7834	-2.2774	-1.5559	-1.2296

Table 3. The LLR distance comparison of proposed method with energy threshold method, blade variance method, spectral subtraction method and noisy signal

Overall Input SNR	LLR				
	Noisy Signal	**Spectral Subtraction**	**Energy Threshold**	**Blade Variance**	**Image Processing**
20dB	1.9613	1.2875	1.3110	2.2015	1.2999
15 dB	2.2331	1.3368	1.3399	2.5084	1.3085
10 dB	2.5290	1.2938	1.2692	2.7826	1.3390
5 dB	2.8526	1.4281	1.4442	3.0311	1.2794
0 dB	3.1859	3.0801	2.9707	3.0550	1.3481
-5 dB	3.4994	3.7833	3.3856	2.6663	1.8707
-10 dB	3.7708	4.0550	3.6413	2.8286	2.4761

The information of tables 2 and 3 has been shown in figures 8 and 9, too. According to these tables and figures, in low SNRs, the proposed method performs better rather than the other methods. In high SNRs, this method has less efficiency. To improve the method in high SNRs, we can use lower values for β.

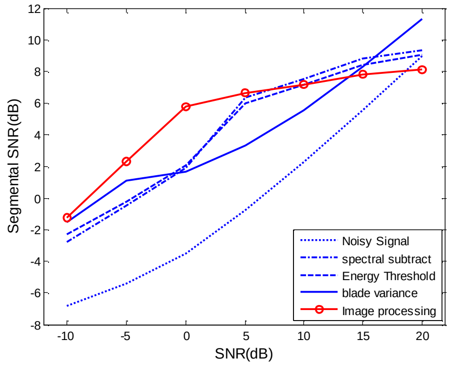


Figure 8. The SegSNR comparison of proposed method with energy threshold method, blade variance method, spectral subtraction method and noisy signal

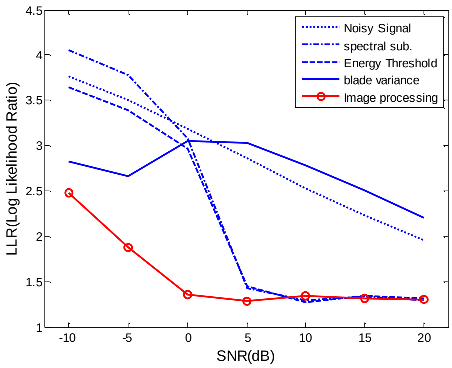


Figure 9. LLR comparison of proposed method with energy threshold method, blade variance method, spectral subtraction method and noisy signal

7. CONCLUSIONS

One of the most important problems in speech enhancement systems is the reduction of musical noise. The musical noise is an undesired effect which reduces the quality of signal and makes it unnatural. In this paper, we have proposed a method that uses the image processing of spectrogram and is based on the symmetry of STFT amplitude around the main frequency of musical noise blocks. Especially in very low SNRs, this method is much superior rather than previous methods; in addition, it requests less computations that makes it applicable for real-time tasks.

ACKNOWLEDGMENT

The authors would like to thank Iran Telecommunication Research Center (ITRC) that funded this research.

REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE *Trans. Acoust., Speech, Signal Processing*, April 1979, vol. ASSP-27, no. 2, pp. 113-120.
[2] W. M. Kushner, V. Goncharoff, C. Wu, and V. Nguyen, "The effects of subtractive type speech enhancement noise reduction algorithms on parameter estimation for improved recognition and coding in high noise environments," in *Proc. of IEEE ICASSP*, 1989, vol. 1, pp. 211–214.
[3] G. Whipple, "Low residual noises peech enhancement utilising time frequency filtering," in *Proc. of IEEE ICASSP*, 1994, vol. 1, pp. 5–8.
[4] Z. Goh, K. Tan, and B.T.G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," *IEEE Trans. on Speech and Audio Processing*, May 1998, vol. 6, pp. 287–292.
[5] S. B. Jebara, "A Perceptual approach to reduce musical noise phenomenon with wiener denoising technique," in *Proc. of IEEE ICASSP*, 2006.
[6] R. J. Niederjohn, P. J. Lee, and F. Josse, "Factors related to spectral subtraction for speech in noise enhancement," in *Proc. of IEEE Int. Conf. on Industrial Electronics and Instrumentation*, pp. 985–996, 1987.
[7] J. H. L. Hansen, and B. Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," in *Proc. of ICSLP*, Dec. 1998.
[8] S. Araki, S. Makino, H. Sawada, and R. Mukai, "Reducing Musical Noise by a fine-shift overlap-add method applied to source separation using a time frequency mask," in *Proc. of IEEE ICASSP*, 2005.