

A Perceptual Approach to Reduce Musical Noise Using Critical Bands Tonality Coefficients and Masking Thresholds

Ch. V. Rama Rao¹, M. B. Rama Murthy², K. Srinivasa Rao³

¹Department of ECE, Gudlavalleru Engineering College, Gudlavalleru, India

²Jayaprakash Narayan College of Engineering, Dharmapur, Mahabubnagar, India

³TRR College of Engineering, Pathancheru, India

Email: chvramaraogec@gmail.com

Received August 18, 2009; revised September 27, 2009; accepted October 19, 2009

Abstract

Traditional noise reduction techniques have the drawback of generating an annoying musical noise. A new scheme for speech enhancement in high noise environment is developed by considering human auditory system masking characteristics. The new scheme considers the masking threshold of both noisy speech and the denoised one, to detect musical noise components. To make them inaudible, they are set under the noise masking threshold. The improved signal is subjected to extensive subjective and objective tests. It is observed that the musical noise is appreciably reduced even at very low signal to noise ratios.

Keywords: Noise Reduction, Musical Noise, Masking Threshold

1. Introduction

In many speech communication systems, enhancing the corrupted speech is a challenging task especially at high noise level. A large number of noise reduction techniques have been proposed in the past. They are based on spectral subtraction [1] and Wiener filtering [2] techniques. The main drawback of these methods is the appearance of an annoying residual noise, often referred to as musical noise. Later techniques developed rely on psychoacoustical considerations. Mainly they exploit the masking properties of the human auditory system. For example according to the enhancement scheme proposed in [3], only audible noise components are estimated and suppressed. Other approaches introduce a perceptual modification on traditional denoising systems [4,5].

In the present paper a new speech enhancement technique is developed for reducing the musical noise. In this work, the auditory masking threshold is estimated for musical noise detection and reduction. Musical noise is detected based on fact that musical noise components present in the enhanced signal lie above the noise masking threshold. On the other hand, the frequency components of noisy speech lie below the noise masking

threshold. Hence, by using some comparison rules musical noise is detected. The detected musical noise components are set under the noise masking threshold and their closet neighbours are smoothed resulting in musical noise reduction.

2. Basic Speech Enhancement System

Let the corrupted speech signal $y(n)$ be represented as

$$y(n) = s(n) + d(n) \quad (1)$$

where $s(n)$ is the clean speech signal and $d(n)$ is the noise signal. The processing is done on a frame-by-frame basis. The Short Time Fourier transform (STFT) is used and the previous model is re-written as

$$Y(m, f) = S(m, f) + D(m, f) \quad (2)$$

where m indicates the frame index and f is the frequency index. The denoised speech short time magnitude $|S(m, f)|$ is obtained using a spectral denoising approach. In this paper, modified Wiener filter [6] is used to denoise the speech signal. The denoised speech is obtained as follows

$$|S(m, f)| = W(m, f) |Y(m, f)| \quad (3)$$

where $W(m, f)$ is the modified Wiener filter gain [6], obtained by including the cross correlation between clean speech signal and noise signal. $W(m, f)$ is given by

$$W(m, f) = \frac{\xi(m, f) + \frac{\delta Y(m, f) D(m, f)}{E\{D^2(m, f)\}}}{\xi(m, f) + 1 + 2 \frac{\delta Y(m, f) D(m, f)}{E\{D^2(m, f)\}}} \quad (4)$$

where $\xi(m, f) = E\{S^2(m, f)\} / E\{D^2(m, f)\}$ is a priori signal to noise ratio (SNR). $\xi(m, f)$ is calculated according to the decision directed approach reported in [7]. δ is the cross correlation coefficient for estimating the correlation between the noisy speech and noise signal in a frame [6]. The modified wiener filter gain function is not only controlled by $\xi(m, f)$ as for conventional Wiener filter but also by δ . When δ is zero noise and clean speech signals are uncorrelated and $W(m, f)$ is reduced to conventional Wiener filter gain function. The proposed approach consists on reducing musical noise existing in denoised speech signal spectrum denoted by $|\hat{s}(m, f)|^2$. The temporal domain enhanced speech is obtained with the following relationship

$$\hat{s}(n) = \text{IFFT} \left[\left| \hat{S}(m, f) \right| e^{j \arg(Y(m, f))} \right] \quad (5)$$

3. Proposed Enhancement Technique

The proposed enhancement technique consists of different steps described below.

- Modified Wiener filter gain function is applied to get denoised speech.
- The noise masking threshold NMT is calculated for both noisy speech and denoised one.
- A musical noise detector is used. For each frequency, it gives a Boolean flag M which indicates the presence or absence of musical noise.
- The musical noise is reduced when present.

3.1. Musical Noise Detection

In order to detect musical noise in denoised speech, perceptual properties of human auditory system are used. There are two steps in detecting musical noise: calculation of noise masking threshold, detection of tonal components in both noisy speech and denoised speech.

3.1.1. Noise Masking Threshold Calculation

The NMT is obtained through modelling the frequency selectivity of the human ear and its masking property. By using masking threshold we distinguish “tone masking noise” and “noise masking tone”. In our context of musical noise detection, we consider only the situation of “noise masking tone”. In fact, the musical noise is a tone signal which is audible during noise components. The NMT is calculated according to principle explained in [8].

3.1.2. Tonal Components Detection

Tonal and non tonal components are identified because their masking models are different. The power spectrum and noise masking threshold of both noisy speech and denoised speech are calculated. Components above noise masking threshold in noisy speech are treated as tonal and belong to speech components. Components above noise masking threshold in denoised speech are marked as tonal and belong to either speech components or musical noise components. Hence, musical noise components can be detected and they are the marked tonal components appearing in denoised speech and not appearing in noisy speech. Figure 1 shows locations of

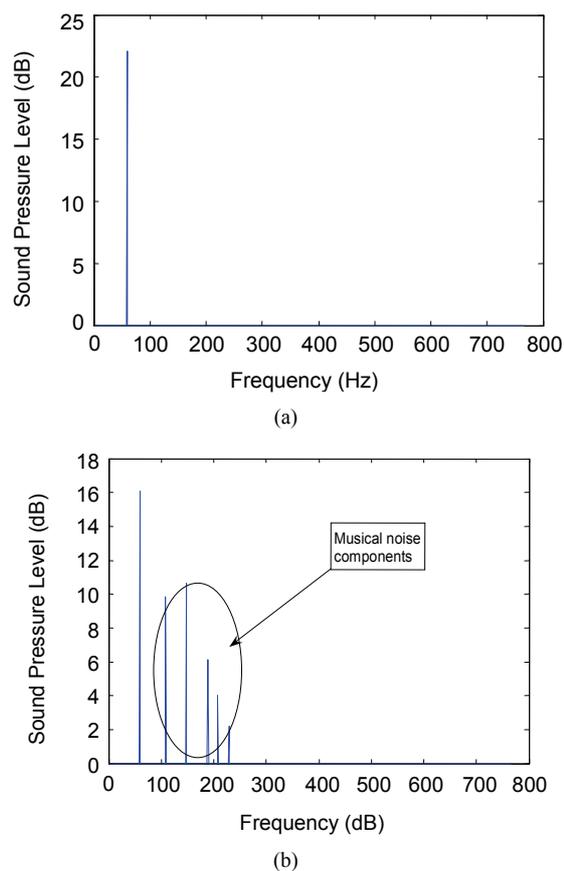


Figure 1. Location of tones in (a) noisy speech (b) denoised speech.

tones in noisy speech and denoised speech. In this work to identify the musical noise tonality coefficient is used. The tonality coefficient α_d is computed for each critical band of denoised speech and α_n for the noisy speech. Musical noise appears in any i^{th} critical band if α_d^i is greater than α_n^i . It becomes audible if the difference $\Delta\alpha_i = \alpha_d^i - \alpha_n^i$ is greater than a certain predetermined threshold τ_i . The threshold τ_i of the i^{th} band depends on critical band order and masking properties of human ear. We are interested in the audibility of tones in the presence of narrow-band noise. A narrow-band noise having 1 bark bandwidth can mask a tone within the same critical band if intensity is below the noise masking threshold NMT_i where the NMT is calculated as follows [5,8]

$$NMT_i = E_i - 5.5 \tag{6}$$

3.1.3. The Experimental Determination of τ_i

The experimental procedure to determine τ_i is as follows:

A white Gaussian noise is considered and power spectrum of each frame is subdivided in critical bands. For each critical band, its energy E_i and its tonality coefficient α_i are computed. For the i^{th} critical band, the power P_i of an additive audible tone which is equal to the noise masking threshold $P_i = NMT_i$ is computed. A sinusoid of the power P_i is injected in the center of the i^{th} critical band and tonality coefficient α_i' is computed. The difference $\alpha_i' - \alpha_i$ represents the threshold τ_i over which an additive tone becomes audible in the presence of narrow-band noise. Experimentally it is observed that τ_i is quite constant for all critical bands and is about $\tau_i=0.06$. Hence in present work τ_i equal to 0.06 is used. Finally, a Boolean flag M , indicating musical noise presence in any critical band is computed using

$$M_i = \begin{cases} 1 & \text{if } \alpha_d - \alpha_n \geq 0.06 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

3.2. Musical Noise Reduction

Musical noise reduction is to remove only the parts responsible of the musical noise character by shifting down the power spectrum of detected musical components under the denoised speech noise masking threshold. In this work correction term, $C(f)$ is used to shift down sufficiently the power spectrum. The estimated power spec-

trum of corrected speech is written as

$$|\hat{S}(m, f)|^2 = \begin{cases} NMT_s(m, f) - C(f) & \text{if } M(m, f) = 1 \\ |\hat{S}(m, f)|^2 & \text{otherwise} \end{cases} \tag{8}$$

where $C(f)$ the correction term is chosen according to subjective listening tests. Values of $C(f)$ for speech and pause frames as given by *Sofia Ben Jebara* [9] indicated in Table 1 are used in the present work. It is observed that the attenuation is small for low frequency and is considerable for high frequency components. During pause, it is constant since distortion and musical component appear in the same way in all frequency bands.

Table1. Correction constants for musical noise reduction.

Frequency band(KHz)		[0,1]	[1,2]	[2,3]	[3,4]
Speech	C(f)	0.5	2	5	10
Pause	C(f)	10	10	10	10

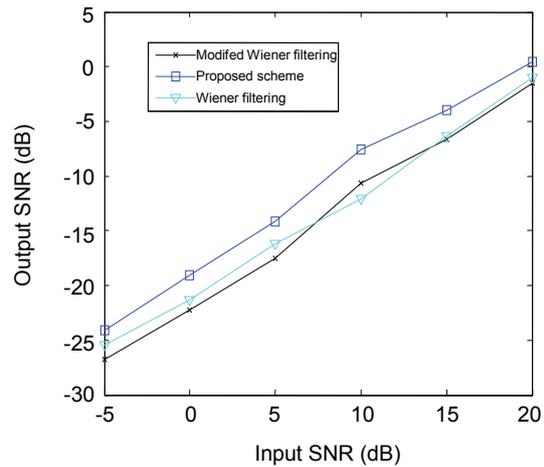


Figure 2. Output segmental SNR values.

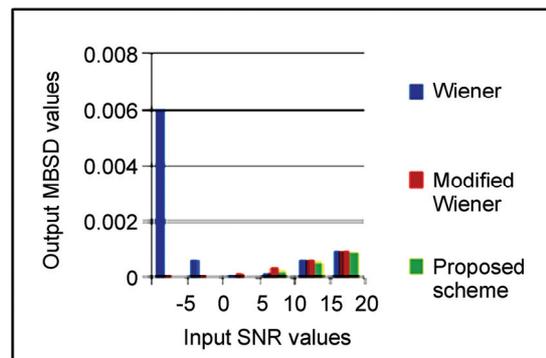


Figure 3. Output MBSD values

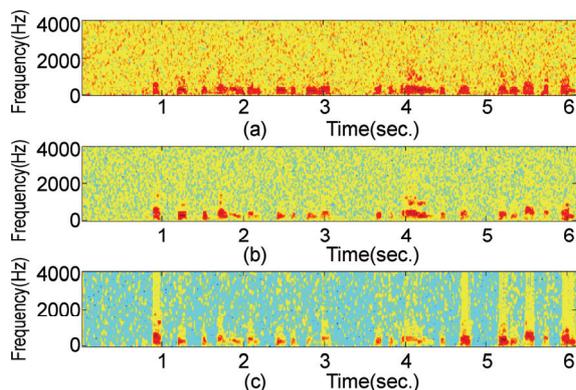


Figure 4. Spectrograms of (a). Noisy speech (b). Denoised speech (c). Enhanced speech by proposed scheme.

4. Results and Discussions

The proposed technique is evaluated using temporal, spectral and perceptual criteria. Segmental signal to noise ratio (SNR_{SEG}) is used as quantitative temporal criteria. For spectral criteria, spectrograms are used and the Modified Bark spectral Distance (MBSD) is used as perceptual criteria [10].

In our simulations, recorded speech samples are used and corrupted with white Gaussian noise and simulations are performed on MATLAB platform. Figure 2 and Figure 3 shows the comparison of performance results of the classical Wiener filtering, modified Wiener filtering and the proposed scheme for different values of signal to noise ratio in terms of SNR_{SEG} and MBSD values respectively. Figure 4 shows spectrogram plots.

Interpretations from the Figures 2, 3 and 4 are as follows:

- The proposed scheme leads to better performance in terms of quality and intelligibility speech signal for all criteria and also it is well noticeable for spectral and perceptual criteria which have good correlation with listening tests.
- Spectrograms are considered in Figure 4. The noisy speech signal is a speech corrupted by a white Gaussian noise whose $SNR=10$ dB. The denoised speech signal by a modified Wiener filtering is affected by a musical noise (isolated points randomly distributed in time and frequency). The amount of such noise is reduced by the proposed scheme.

5. Conclusions

In this work, a new enhancement scheme for reducing

musical noise imposed by a modified Wiener filtering is proposed. The masking characteristics of the human ear are used to detect and to reduce musical noise. Simulation results show that this scheme provides better results in terms of temporal, spectral and perceptual criteria.

6. References

- [1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transaction Acoustics, Speech and Signal Processing*, Vol. ASSP-27, No. 2, pp. 113–120, April 1979.
- [2] Y. Ephraim and D. Mallah, "Spectral enhancement using optimal non-linear spectral amplitude estimation," on *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 1118–1121, 1983.
- [3] A. Akbari-Azrani, R. Le bouquin Jannes, and G. Faucon, "Optimizing speech enhancement by exploiting masking properties of the human ear," on *Proceedings International Conferences on Acoustic, Speech and Signal Processing ICASSP*, IEEE, pp. 800–803, 1995.
- [4] N. Virag, "Single channel speech enhancement based on masking properties of human auditory system," *IEEE Transactions on Speech and Audio Processing*, Vol. 7, No. 2, pp. 126–137, February 1999.
- [5] K. A. Sheela, CH. V. R. Rao, K. S. Prasad, and A. V. N. Tilak, "A new noise reduction pre-processor for mobile voice communication using perceptually weighted spectral subtraction method," *3rd International Conferences on Mobile Ubiquitous and Pervasive Computing*, VIT University, 16-19 December 2006.
- [6] CH. V. R. Rao, M. B. R. Murthy, and K. S. Rao, "Speech enhancement using modified Wiener filter," *National Conference on Futuristic Advancements in Computing & Electronics*, Deccan College of Engineering & Technology, 19-21 March, 2009.
- [7] Y. Ephraim and D. Mallah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Transaction on Speech Audio Processing*, Vol. ASSP-32, pp. 1109–1121, 1984.
- [8] J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," *IEEE, Journal on Selected Areas of Communication*, Vol. 6, pp. 314–323, 1988.
- [9] S. B. Jebara, "A perceptual approach to reduce musical noise phenomenon with wiener denoising technique," *proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2006.
- [10] W. Yan, M. Dixon, and R. Yantorno, "A modified bark spectral distortion measure which uses noise masking threshold," on *Proceedings of the Speech Coding Workshop IEEE*, pp. 55–56, 1997.