An Enhanced Clarity of Husky Voice by Dissonant Frequency Filtering

Sangki Kang* · Seong-Joon Baek**

ABSTRACT

There have been numerous studies on the enhancement of noisy speech signal. In this paper, we propose a new speech enhancement method, that is, a filtering of a dissonant frequency combined with noise suppression algorithm. The simulation results indicate that the proposed method provides a significant gain in voice clarity. Therefore if the proposed enhancement scheme is used as a pre-filter, the perceptual clarity of husky voice is greatly enhanced.

Keywords: Dissonant Frequency Filtering, Noise Suppressor

1. Introduction

Degradation of the quality or intelligibility of speech caused by the acoustic background noise is common in most practical situations. In general, the addition of noise reduces intelligibility and degrades the performance of digital voice processors used for applications such as speech compression and recognition. Therefore, the problem of removing the uncorrelated noise component from the noisy speech signal, i.e., speech enhancement, has received considerable attention. There have been numerous studies on the enhancement of the noisy speech signal. Many different types of speech enhancement algorithms have been proposed and tested[1]-[7]. They can be grouped into three categories. The first category contains speech enhancement algorithms based on the short-time spectral estimation such as the spectrum subtraction[1] and Wiener filtering[2][3] techniques. The algorithms in the second category are comb filtering and an adaptive noise canceling technique which exploit the quasi-periodic nature of the speech signal [4][5][6]. The third category contains algorithms that are based on the statistical model of the speech signal and use hidden Markov model (HMM) or expectation and maximization algorithm[7].

In this paper, a new speech enhancement scheme using a filtering of a dissonant frequency (especially *C#* and *F#* in each octave band when reference frequency is *C*) combined with noise

* SW Lab. Telecommunication R&D Center Samsung Electronics Co., Ltd
** Corresponding author, The Dept. of Electronics Engr., Chonnam National Univ.

suppression algorithm[9] is proposed.

2. A Dissonant Frequency Filtering Combined With Noise Suppression

The standard musicological definition is that a musical interval is consonant if it sounds pleasant or restful. Dissonance, on the other hand, is the degree to which an interval sounds unpleasant or rough. Dissonant intervals generally feel tense and unresolved. When reference frequency is *C*, *C#* and *F#* are such examples. Among them, *F#* is known as "the Devil's interval" in music[10]. So elimination of *C#* and *F#* would help speech less annoying and more pleasant.

In a DFF(Dissonant Frequency Filtering) combined with noise suppression (NS) algorithm to enhance perceptual voice clarity, it is important to accurately estimate the fundamental frequency that is closely related to a dissonant frequency filtering. A filtering of a dissonant frequency combined with noise suppression algorithm based on the improved fundamental frequency estimation [8] which is developed in frequency domain is as follows.

First, the noisy input signal is processed using noise suppression algorithm. Second, frequency transform of the signal using Fast Fourier Transform (FFT) is carried out and then an improved fundamental frequency estimation is performed [8]. Next, filtering of the dissonant frequency based on the obtained fundamental frequency (pitch) is performed.

The dissonant frequency F_{d1} and F_{d2} corresponding to *C#* and *F#* relative to fundamental frequency are defined as follow [11].

$$F_{d1} = F_0 \times 2^{(n+1/12)}, \quad n = 0, 1, \cdots, 7 \tag{1}$$

$$F_{d2} = F_0 \times 2^{(n+6/12)}, \quad n = 0, 1, \cdots, 7 \tag{2}$$

Here F_{d1} and F_{d2} are dissonant frequencies and F_0 is a fundamental frequency estimated by the improved fundamental frequency estimation using parametric cubic convolution [8] and n which represent octave band varies from 0 to 7 on condition that F_{d1} and F_{d2} is less than the half of sampling frequency. The dissonant frequency F_{d1} and F_{d2} are filtered out if the following condition is satisfied.

$$F_0 \times 2^{(n+1/24)} < F_{d1} < F_0 \times 2^{(n+3/24)}, \quad n = 0, 1, \cdots, 7 \tag{3}$$

$$F_0 \times 2^{(n+11/24)} < F_{d2} < F_0 \times 2^{(n+13/24)}, \quad n = 0, 1, \cdots, 7 \tag{4}$$

where each range corresponds to a half tone or a semitone.

A simple filtering algorithm is adopted in this paper. If speech shows a peak in the region defined above, that peak is smeared out, that is, the magnitude in that region is lowered to the neighboring magnitude while the phase is kept. Finally, the inverse Fourier Transform of filtered signal is carried out. This procedure is illustrated in Figure 1.

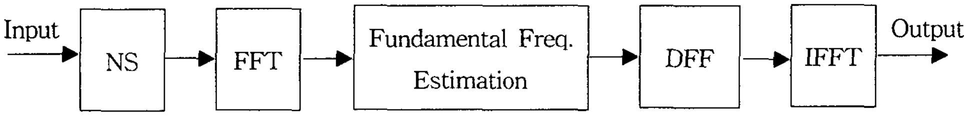

Figure 1. Voice clarity improvement scheme.

3. Experimental Results

The database for more reliable performance evaluation consists of 30 speech files collected from 10 speakers (6 males and 4 females) whose intelligibility is worse than normal speakers, each one delivering 3 Korean sentences. All utterances were sampled at 8 *kHz* with 16-bit resolution and 1024-point FFT is used. Noise types considered in our experiments are white Gaussian noise, babble noise and car noise recorded inside a car moving approximately at a speed of 80 km/h. We obtained the noisy speech by adding noise to clean speech with the noise power being adjusted to achieve SNR = 5, 10, and 15 dB. The noise suppression algorithm in IS-127 [9] is adopted. In order to evaluate the performance of the proposed enhancement scheme, MOS (mean opinion score) tests were conducted. 20 listeners participated in the test. During the testing period, each individual used a high-quality handset as a listening device in a quiet room. The recordings in each presentation were played randomly. The listeners were asked to give a score ranging from 1 (Bad) to 5 (Excellent) according to the perceived quality. The simulation results are represented in Table 1 and Figure 1. The results clearly indicate that the proposed method provides a significant gain especially in a husky voice corrupted with babble noise.

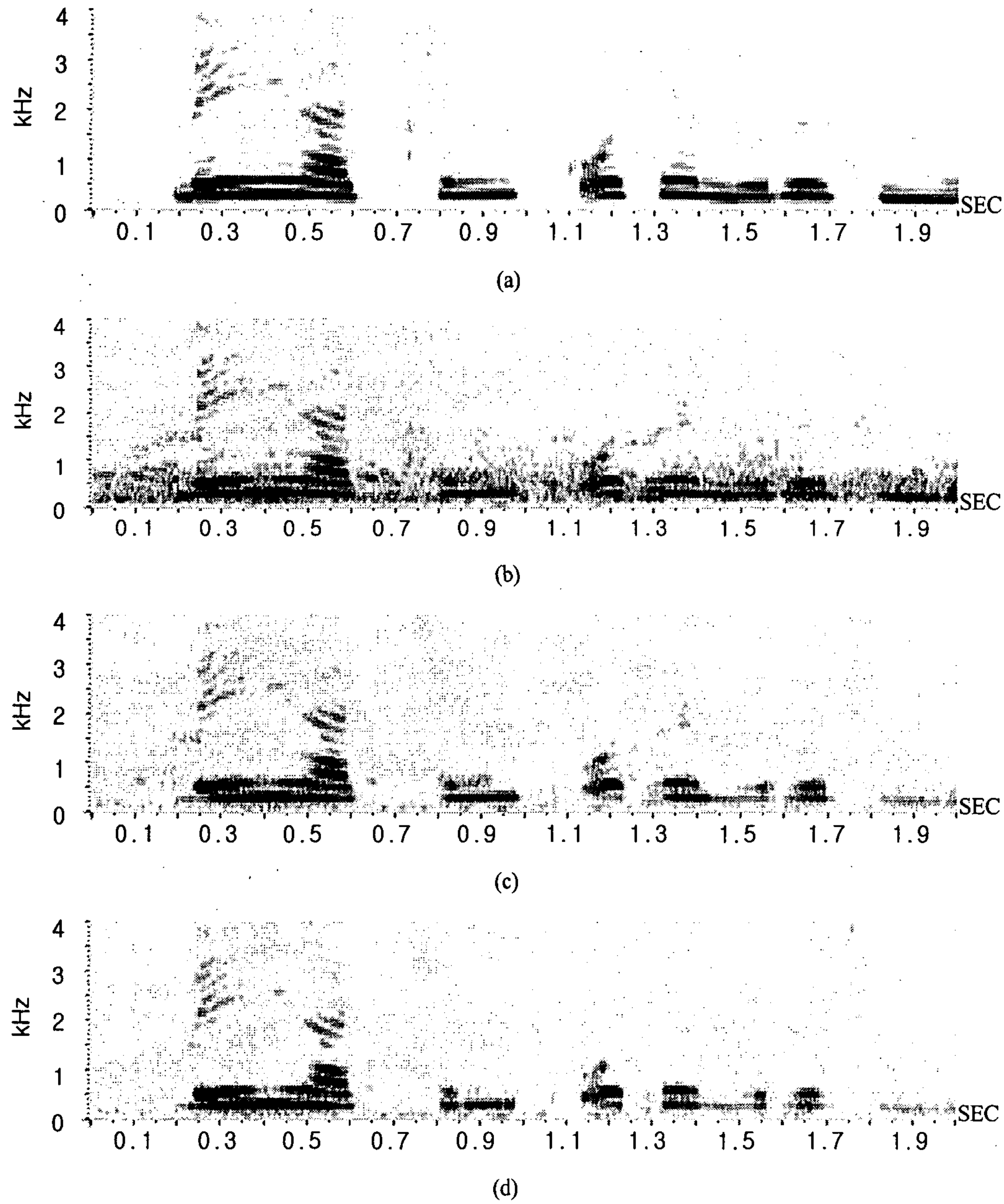

Figure 2. Spectrograms of speech
(a) Clean signal (b) Noisy signal (c) NS (d) NS + DFF

Table 1. Results of subjective measures of enhanced voice clarity without and with noise suppressor

Noise Types	SNR	MOS		
		Unprocessed data	Processed data using DFF only	Processed data using DFF and NS
White Gaussian Noise	5dB	2.13	2.25	2.87
	10dB	2.25	2.47	2.97
	15dB	2.49	2.66	3.18
Babble Noise	5dB	2.11	2.43	2.99
	10dB	2.36	2.66	3.27
	15dB	2.54	2.87	3.56
Car Noise	5dB	2.29	2.45	2.85
	10dB	2.33	2.57	3.16
	15dB	2.51	2.66	3.37

4. Conclusions

We proposed a new speech enhancement scheme, that is, a filtering of a dissonant frequency combined with noise suppression algorithm. The simulations results indicate that the proposed method delivered improvement in terms of both speech intelligibility and perceived clarity. Therefore if the proposed enhancement scheme is used as a pre-filter, the perceptual quality of voice clarity is greatly enhanced.

References

[1] Boll, S. F. 1979. "Suppression of acoustic noise in speech using spectral subtraction." *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-29, pp. 113-120.

[2] Lim, J. S. & Oppenheim, A. V. 1978. "All-pole modeling of degraded speech." *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, pp. 197-210.

[3] Hansen, J. & Clements, M. 1991. "Constrained iterative speech enhancement with application to speech recognition." *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-39, No.4, pp. 21-27.

[4] Samber, M. R. 1978. "Adaptive noise canceling for speech signals." *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, pp. 419-423.

[5] Lim, J. S., Oppenheim, A. V. & Braida, L. D. 1991. "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition." *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, No. 4, pp. 354-358.

[6] Boll, S. F. & Pulsipher, D. C. 1980. "Suppression of acoustic noise in speech using two microphone adaptive noise cancelation." *IEEE Trans. Acoust., Speech, Signal Processing*,

Vol. ASSP-28, No. 6, pp. 752-753.
[7] Ephraim, Y. 1992. "Statistical-model-based speech enhancement systems." *Proc. of IEEE*, Vol. 80, No. 10, pp. 1526-1555.
[8] Pang, Hee-Suk., Baek, SeongJoon. & Sung, Koeng-Mo. 2000. "Improved fundamental frequency estimation using parametric cubic convolution." *IEICE Trans. Fundamentals*, Vol. E83-A, No. 12, pp. 2747-2750.
[9] TIA/EIA/IS-127. 1997. "Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems."
[10] Tennenbaum, J. 1991-92. "The foundations of scientific Musical Tunning." *FIDELIO Magazine*, Vol. 1 No. 1.
[11] Howard, D. M. & Angus, J. 1996. *Acoustics and Psychoacoustics*, Focal Press.

received: November 09, 2005
accepted: December 10, 2005

▲ Sangki Kang
SW Lab. Telecommunication R&D Center Samsung Electronics Co., Ltd
Dong Suwon P.O.BOX 105, 416, Maetan-3dong, Paldal-gu, Suwon-si,
Gyeonggi-do, Korea 442-600
Tel: +82-31-279-5486 Fax: +82-31-279-4606
E-mail: sangki.kang@samsung.com

▲ Seong-Joon Baek
The Department of Electronics Engineering
Chonnam National University, Gwangju, Korea
Tel: +82-62-530-1795 Fax: +82-62-530-1759
E-mail: tozero@chonnam.ac.kr