THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.23, NO.1E 2004. 3 pp. 16~18

Filtering of a Dissonant Frequency Combined with Noise Reduction for Speech Enhancement

Sangki Kang*, Youn-Jeong Lee**, Ki Yong Lee**
*Modem System Lab. Telecommunication R&D Center Telecommunication Network Samsung Electronics Co., Ltd.
**School of Electronic Engineering, Soongsil University
(Received February 12 2004; revised February 23 2004; accepted March 15 2004)

Abstract

There have been numerous studies on the enhancement of the noisy speech signal. In this paper, I propose a completely new speech enhancement method, that is, a filtering of a dissonant frequency combined with noise reduction algorithm. The simulation results indicate that the proposed method provides a significant gain in audible improvement compared with the conventional method. Therefore if the proposed enhancement scheme is used as a pre-filter, the perceptual quality of speech is greatly enhanced.

***Keywords**: Dissonant frequency filtering (DFF), Noise reduction, Fundamental frequency estimation*

I. 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[4-6]. 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 (EM)[7].

In this paper, a completely new speech enhancement scheme using a filtering of a dissonant frequency (especially C#, B and F# in each octave band when reference frequency is C) combined with noise reduction algorithm [9] is proposed.

Corresponding author: Sangki Kang (sangi.kang@samsung.com)
Modem System Lab. Telecommunication R&D Center Samsung Electronics Co., LTD 416, Maetan-3dong, Yeongtong-gu, Suwon-city, Gyeonggi-do, Korea 442-600

II. A Dissonant Frequency Filtering Combined with Noise Reduction

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#, B and F# are such examples. Among them, F# is known as "the Devil's interval" in music[10]. So C#, B and F# are eliminated to make speech less annoying and more pleasant.

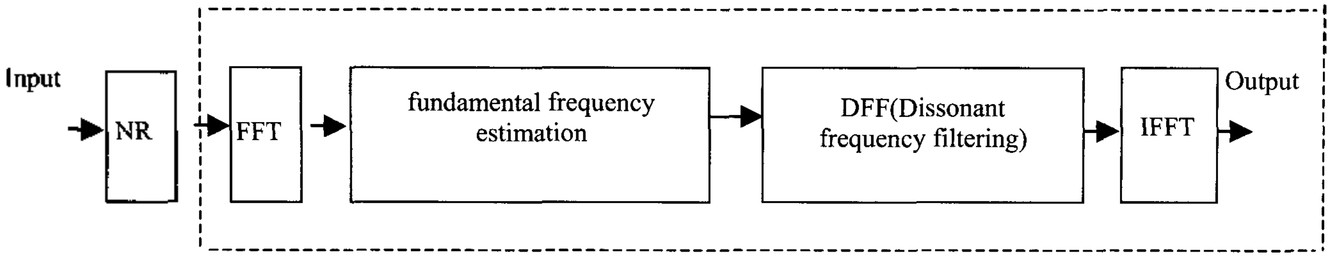

Fig.1. Speech quality improvement scheme.

In a dissonant frequency filtering combined with noise reduction algorithm to enhance perceptual speech intelligibility, it is important to accurately estimate the fundamental frequency that is closely related to a dissonant frequency filtering. As follows is a filtering of a dissonant frequency combined with noise reduction algorithm based on the improved fundamental frequency estimation[8] which is developed in frequency domain.

First, the noisy input signal is processed using noise reduction 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} , F_{d2} and F_{d3} corresponding to C#, B and F# relative to fundamental frequency are defined as follow [11].

$$F_{d1} = F_0 \times 2^{(n+1/12)}, \qquad n = 0,1,\ldots,7 \qquad (1)$$
$$F_{d2} = F_0 \times 2^{(n+6/12)}, \qquad n = 0,1,\ldots,7 \qquad (2)$$
$$F_{d3} = F_0 \times 2^{(n+11/12)}, \qquad n = 0,1,\ldots,7 \qquad (3)$$

Here F_{d1}, F_{d2} and F_{d3} are dissonant frequency, F_0 is fundamental frequency obtained 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}, F_{d2} and F_{d3} is less than the half of sampling frequency. The dissonant frequency F_{d1} , F_{d2} and F_{d3} are filtered out if the following condition is satisfied.

$$F_0 \times 2^{(n+1/24)} < F_{d1} < F_0 \times 2^{(n+3/24)}, \quad n = 0,1,\ldots,7 \qquad (4)$$
$$F_0 \times 2^{(n+11/24)} < F_{d2} < F_0 \times 2^{(n+13/24)}, \quad n = 0,1,\ldots,7 \qquad (5)$$
$$F_0 \times 2^{(n+21/24)} < F_{d3} < F_0 \times 2^{(n+23/24)}, \quad n = 0,1,\ldots,7 \qquad (6)$$

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

A simple filtering algorithm is employed 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 is illustrated in Fig.1.

III. Experimental Results

The database for performance evaluation consists of ten speech files collected from five speakers (three males and two females), each one delivering two Korean sentences. Also we obtained six husky voice files from three speakers (two males and one female) whose intelligibility is worse than normal speakers. All utterances were sampled at 8kHz with 16-bit resolution. Also 256-point FFT is used. Noise types considered in our experiments include white Gaussian noise, babble noise and car noise recorded inside a car moving approximately at a speed of 80 km/h. I obtained the noisy speech by adding noise to clean speech with the noise power being adjusted to achieve SNRs=5, 10, and 15 dB. The noise reduction algorithm in IS-127 [9] for NR is adopted. In order to evaluate the performance of the proposed enhancement scheme, MOS (mean opinion score) tests were conducted. Fifteen 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 results of subjective measures of enhanced speech without Noise Reduction are also added in Table.1.

Table 1. Results of subjective measures of enhanced speech without and with Noise Reduction.

Types	SNR	MOS			
		Unprocessed	DFF(processed)	NR (IS-127)	NR + DFF (Proposed)
White Gaussian noise	5dB 10dB 15dB	2.13 2.25 2.49	2.67 2.79 3.00	2.82 2.87 3.01	2.97 3.08 3.21
Babble Noise	5dB 10dB 15dB	2.11 2.36 2.54	2.92 3.01 3.10	2.91 2.98 3.06	3.18 3.33 3.53
Car Noise (motor noise)	5dB 10dB 15dB	2.29 2.33 2.51	2.75 2.84 3.19	2.88 2.97 3.21	3.02 3.19 3.34
Husky Voice (white noise)	5dB 10dB 15dB	2.03 2.15 2.39	2.65 2.81 2.92	2.67 2.78 2.98	3.02 3.27 3.45

The simulation results indicate that the proposed method provides a significant gain in audible improvement especially for a husky voice signal and speech signal corrupted by babble noise.

IV. Conclusions

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

References

1. S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29, pp.113-120, April 1979.
2. J.S. Lim and A.V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-26, pp.197-210, June 1978.
3. J. Hansen and M. Clements, "Constrained iterative speech enhancement with application to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-39(4), pp.21-27, April 1991.
4. M.R. Samber, "Adaptive noise canceling for speech signals," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-26, pp.419-423, Oct. 1978.
5. J.S. Lim, A.V. Oppenheim, and L.D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-26(4), pp.354-358, April 1991.
6. S.F. Boll and D.C. Pulsipher, "Suppression of acoustic noise in speech using two microphone adaptive noise cancelation," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-28(6), pp.752-753, Dec. 1980.
7. Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proc. of IEEE*, 80(10), pp.1526-1555, Oct. 1992.
8. Hee-Suk Pang, SeongJoon Baek, Koeng-Mo Sung, "Improved fundamental frequency estimation using parametric cubic convolution," *IEICE Trans. Fundamentals*, E83-A(12), pp. 2747-2750, 2000.
9. TIA/EIA/IS-127, "Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems," 1997.
10. Jonathan Tennenbaum, "The foundations of scientific Musical Tunning," *FIDELIO Magazine*, 1(1), Winter 1991-92.
11. David M. Howard and James Angus, "*Acoustics and Psychoacoustics*," Focal Press 1996.

[Profile]

● Sangki Kang

The Journal of the Acoustical Society of Korea, Vol. 22, NO.3E

● Youn-Jeong Lee

The Journal of the Acoustical Society of Korea, Vol. 22, NO.7

● Ki Yong Lee

The Journal of the Acoustical Society of Korea, Vol. 15, NO.3E