Simulation of Spectral Subtraction Based Noise Reduction Method

Zhixin Chen

ILX Lightwave Corporation
Bozeman, Montana, USA
chen.zhixin.mt@gmail.com

Abstract — **Noise reduction is a very meaningful but difficult task and it has been a subject of intense research in recent years. This paper introduces two popular noise reduction techniques and presents our simulation result of a noise reduction system. It is shown that the system reduces the noise almost completely while keeps the enhanced speech signal very similar to the original speech signal.**

Keywords - Embedded Systems; Digital Signal Processing; Noise Reduction.

I. INTRODUCTION

Noise is one kind of sound that is unexpected or undesired [1, 10]. It can be divided into non-additive noise and additive noise. The non-additive noise includes multiplier noise and convolutional noise, which can be transformed into additive noise through homomorphic transform. The additive noise includes periodical noise, pulse noise, and broadband noise [2, 4]. The noise generated by the engine is one kind of periodical noise while the one generated from explosion, bump, or discharge is pulse noise. There are many kinds of broadband noise, which may include heat noise, wind noise, quantization noise, and all kinds of random noise such as white noise and pink noise.

Loud and persistent noise can be very dangerous to human's health since it influences the function of the human body system and thus causes the body to respond in ways that may lead to stress disorder, irritability, and sleepiness [5]. Noise with sufficient intensity can temporally or even permanently damage hearing.

Consequently, noise has gradually become a new dangerous pollutant. Fortunately, people start to notice this and they have taken an action against excessive noise in recent years [7].

The ideal way to treat the noise pollution would be just to get rid of it. This can be done for environmental noise like traffic noise or machinery noise by using a barrier material or a passive absorber [1]. However, in most cases, a noise polluted signal is more complicated because it is desirable to cancel the noise completely but retain the desired signal. An example is the detection of a weak signal in a noisy environment or a telephone interview in the noisy environments such as cars, airports and laboratories [8]. We can imagine the sad expression of the interviewee if we design a system to get rid of all the sound, which includes the noise and his conversation with the interviewer. The noise reduction is a very meaningful but difficult task and it has been a subject of intense research in recent years.

This paper introduces two popular noise reduction techniques and presents our simulation result of a noise reduction system.

The remaining part of this paper is organized as follows. First, the popular noise reduction techniques, including spectral subtraction and Wiener filtering, are reviewed. Then, the simulation of a spectral subtraction based noise reduction system is presented. Finally, this paper is concluded with a summary of results.

II. NOISE REDUCTION TECHNIQUES

In a phone interview in a noisy environment, the speech signal may be distorted by the background noise, which can be generated by the machine, computer, or even the electronic fans. If a handsfree telephone is used in the interview, the intensity of the background noise may be even stronger than the speech signal. The noise will thus distort the speech and make it hardly intelligible. In order to improve the intelligibility, the noise needs to be attenuated to enhance the speech signal. Figure 1 shows the block diagram of a noise reduction system. In this figure, the noisy speech signal $X(n)$ is the combination of the original speech signal $S(n)$ and the noise $N(n)$. The noisy speech signal $X(n)$ passes through a noise reduction system to get a clean speech signal $Y(n)$, which is similar to the original speech signal $S(n)$.
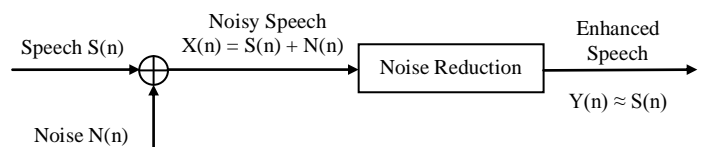

Figure 1: The block diagram for noise reduction system

The intelligibility and naturalness of the enhanced signal, improvement of signal-to-noise ratio, signal delay, and computational complexity are four key criteria for the performance of a noise reduction system [9]. It is obvious that after noise reduction, a high quality speech with high intelligibility and naturalness is desired. At the same time, the interference needs to be as weak as possible, which also means that a high signal-to-noise ratio is needed. It is reported that if the delay for the telephone communication is larger than 100 milliseconds, the delay can be perceived [3, 6]. Consequently,

short signal delay and low computational complexity are also desired in the noise reduction systems.

Noise reduction algorithms generally operate in the frequency domain, which includes adaptive filtration, autocorrelation methods, methods based on parametric models of spectrum estimation, and techniques employing intelligent algorithms [9, 11]. The existing noise reduction systems are mainly based on spectral subtraction and Wiener filtering.

A. Spectral Subtraction

Spectral subtraction is the most popular noise reduction method [2, 9, 11]. This method operates in the frequency domain and assumes that the spectrum of the input noisy signal can be expressed as the sum of the speech spectrum and the noise spectrum. Figure 2 shows the block diagram for the spectral subtraction method. The noise spectrum is first estimated and then subtracted from the noisy speech spectrum to get the clean speech spectrum.
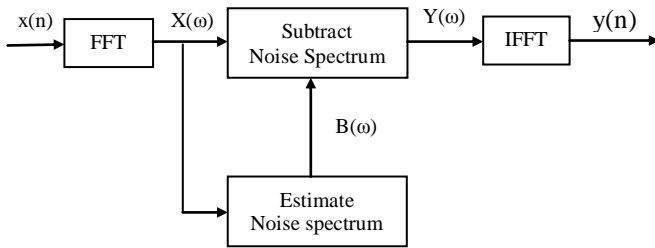

Figure 2: the block diagram for the spectral subtraction method

In Figure 2, it is assumed that x(n) is a discrete-time noisy sequence with $x(n) = s(n) + b(n)$, where s(n) is the desired signal and b(n) is the unwanted background noise. Here s(n) and b(n) are assumed to be wise-sense stationary and uncorrelated random processes with power spectral density functions denoted by $S_s(\omega)$ and $S_b(\omega)$, respectively. The sum of the power spectra $S_x(\omega) = S_s(\omega) + S_b(\omega)$ is used to recover the desired signal s(n). Although speech signal is time-variant but it can be considered stationary in a short time interval. The short time Fourier transform (STFT) analysis can be used in the short-time segments. In the time domain, $x_{pL}(n) = w[pL - n](s(n) + b(n))$, where L is frame length and p is an integer. The frequency domain expression is $X(pL, \omega) = S(pL, \omega) + B(pL, \omega)$, where $S(pL, \omega)$, $B(pL, \omega)$ and $X(pL, \omega)$ are the STFTs of the object s(n), the background noise b(n), and the measurement x(n) computed at frame interval L, respectively. So, the STFT magnitude squared of x(n) can be written as:

$$|X(pL, \omega)|^2 = |S(pL, \omega)|^2 + |B(pL, \omega)|^2 + S^*(pL, \omega)B(pL, \omega) + S(pL, \omega)B^*(pL, \omega)$$

(1)

Here the objective is to obtain an estimate of $|S(pL, \omega)|^2$.

In this signal estimation approach, the STFT phase is not estimated. Consequently, the best for each short-time segment is an estimate of the form $\hat{S}(pL, \omega) = |S(pL, \omega)| e^{j\angle X(pL, \omega)}$. This means that the ideal STFT estimate consists of the clean STFT magnitude and noisy measure STFT phase. This is referred as the theoretical limit in estimating the original STFT when only the STFT magnitude is estimated. By considering the threshold of perception of phase deviation due to additive noise, it has been shown that speech degradation is not perceived with an average short-time segmental SNR greater than 6dB for the theoretical limit in the equation above-mentioned. However, when this SNR is considerably below 6 dB, a roughness of the reconstruction will be perceived.

Assume that an estimate of the power spectrum of the noise, denoted by $\hat{S}_b(\omega)$, which is typically obtained by averaging over multiple frames of a known noise segment, is given. Also assume that the noise and object sequences are uncorrelated. Then with short-time analysis, an estimate of the object's short-time squared spectral magnitude is expressed as

$$|\hat{S}(pL, \omega)|^2 = \begin{cases} |X(pL, \omega)|^2 - \hat{S}_b(\omega) & if \quad |Y(pL, \omega)|^2 - \hat{S}_b(\omega) \geq 0 \\ 0 & otherwise \end{cases}$$

(2)

When this magnitude estimate is combined with the measured

phase, the STFT estimate $\hat{S}(pL, \omega) = |\hat{S}(pL, \omega)| e^{j\angle X(pL, \omega)}$ is obtained. An objected signal estimate can then be formed with overlap-add (OLA), filter-bank summation (FBS), or least-squared-error (LSE) synthesis [2, 11].

B. Wiener Filtering

An alternative way to spectral substraction for recovering an object suquence s(n) from a sequence $x(n) = s(n) + b(n)$ is to find a linear filter h(n) such that the sequence $\hat{s}(n) = x(n) * h(n)$ minimizes the expected value of $|s(n) - \hat{s}(n)|^2$ [7]. Under the condition that the signals s(n) and b(n) are uncorrelated and stationary, the frequency-domain solution to this stochastic optimization problem is given by the suppression filter

$$H_s(\omega) = \frac{S_s(\omega)}{S_s(\omega) + S_b(\omega)},$$

(3)

which is referred to as the Wiener filter [9, 11].

When the signals s(n) and b(n) are uncorrelated and stationary, the Wiener filter provides noise suppression without considerable distortion in the estimated object. The required power spectra, $S_s(\omega)$ and $S_b(\omega)$, can be estimated by averaging over multiple frames when sample functions of s(n) and b(n) are provided. But the desired signal and background are typically nonstationary in the sense that their power spectra change over time, which also means that they can be expressed as time-varying functions $S_s(n, \omega)$ and $S_b(n, \omega)$. Consequently, each frame of the STFT is processed by a different Wiener filter. For the simplified case of a stationary background, the time-varying Wiener Filtering can be expressed as:

$$H_s(pL,\omega) = \frac{\hat{S}_s(pL,\omega)}{\hat{S}_s(pL,\omega) + \hat{S}_b(\omega)},$$

(4)

where $\hat{S}_s(pL,\omega)$ is an estimate of $S_s(n,\omega)$ on each frame and $\hat{S}_b(\omega)$ is an estimate of $S_b(\omega)$. As with spectral subtraction, for the Wiener filtering, an enhanced waveform is recovered from the modified STFT, $\hat{S}(pL,\omega) = Y(pL,\omega)H_s(pL,\omega)$, by overlap-add (OLA), filter-bank summation (FBS), or least-squared-error (LSE) synthesis.

III. SIMULATION OF NOISE REDUCTION

One thing for both spectral subtraction and Wiener filtering is how to estimate the noise. According to [9, 11], there are two main methods. The simpler form is the analysis during speech pauses while the more complicated one is a minimum statistics algorithm. Another thing for spectral subtraction and Wiener filtering is that their processed output signals suffer from musical noise. These artifacts are due to randomly distributed spectral peaks in the residual noise spectrum. In [12], Ephraim and Malah proposed two popular methods to reduce the musical noise phenomenon.

In this section, the simulation result of a noise reduction system is given. This system is based on spectral subtraction. It processes the noisy speech signal with a sampling frequency of 8 KHz or 16 KHz. It first transforms the signal from time domain to frequency domain and then divides the frequency domain into 16 or 19 channels, depending on sampling frequency. Then it estimates the noise signal during speech pause and subtracts the noise from the noisy speech in every channel. Finally, it sums the signal from every channel and then converts the signal from the time domain back to frequency domain. This algorithm performs well in modest noisy environment with low computational complexity. The subjective reconstructed speech quality is good. Figure 3 lists the waveform and spectrogram of the noisy speech and reconstructed speech. It can be seen from both the waveform and the spectrogram that this system reduces the noise almost completely while keeps a speech signal very similar to original speech signal.
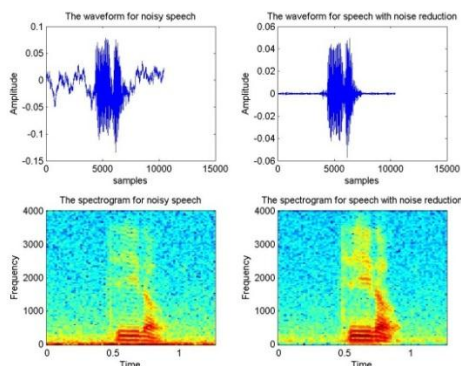

Figure 3: The waveform and spectrogram of the noisy speech and enhanced speech.

IV. CONCLUSION

This paper introduced the definition, classification, and the hazard of noise. It then described two main techniques, spectral subtraction and Wiener filtering, for noise reduction. The paper also presented the simulation result for a noise reduction system based on spectral subtraction.

Currently, there is a lot of research on the new techniques for noise reduction. First, the spectral subtraction and Wiener filtering did not rely on a speech model. Researchers found that a noise reduction filter that exploits estimated speech model parameters can be designed in the noise reduction system.

For example, the Wiener filter can be constructed with an object power spectrum estimation that is based on an all-pole vocal tract transfer function. This filter can then be applied to enhance speech, just like the nonparametic case.

Second, in the phenomenon of auditory masking, one sound component is concealed by the presence of another sound component. Hence, the auditory masking principle can be used in reducing the perception of noise.

Finally, the wavelet transform can also be used to replace the STFT method for noise reduction.

REFERENCES

[1] Acoustical Solutions, Inc. "Introduction to Noise Control", Acoustic Education.

[2] Yang Xingjun, Chi Huisheng, "Digital Speech Signal Processing", 1992 (Chinese).

[3] Zhixin Chen, "Design and implementation on a sub-band based acoustic echo cancellation approach," International Journal of Advanced Computer Science and Applications, vol. 2, issue 6, 2011.

[4] Kinsler, Lawrence etc., "Fundamentals of Acoustics", 4th ed., Wiley & Sons, 1999.

[5] The Kansas City Health Department, "Noise Pollution".

[6] Zhixin Chen, "Investigation on simulation and measurement of reverberation for small rooms," International Journal of Advanced Computer Science and Applications, vol. 2, issue 7, 2011.

[7] "Introduction to Noise Control", http://www.polytechinc.com/noise.htm

[8] Joachim Holzfuss, "Active Noise Reduction", 1997.

[9] Stefan Schmitt, Malte Sandrock, "Single Channel Noise Reduction Algorithm for Handsfree Operation in Distorted Enviroments", 2001.

[10] Zhixin Chen and Robert C. Maher, "Analytical expression for impulse response between two nodes in 2-D rectangular digital waveguide mesh," IEEE Signal Processing Letters, vol. 15, pp. 221-224, 2008.

[11] Thomas F. Quatieri, "Discrete-Time Speech Signal Processing", Prentice Hall Signal Processing Seriers.

[12] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Transactions on Acoustics, Speech, and Signal processing, vol. 32, no. 6, pp. 1109-1121, December 1984

AUTHORS PROFILE

Zhixin Chen is a Firmware Engineer in ILX Lightwave Corporation. He holds a BS degree from Xiamen University in China, a MS degree from Xiamen University in China, and a Ph.D. degree from Montana State University in USA, all in Electrical and Computer Engineering.

His research interest and working experience are in the area of acoustics, audio, and speech processing, multimedia communication, and embedded system design for high power current source, temperature controller, and optical power meter.