NOISE REDUCTION BASED ON CROSS TF $\varepsilon$-FILTER

Tomomi Abe

Major in Pure and Applied Physics, Waseda university 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

Mitsuharu Matsumoto, Shuji Hashimoto

Department of Applied Physics, Waseda university 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

Keywords: Noise reduction, speech enhancement, $\varepsilon$-filter, time-frequency domain.

Abstract: A time-frequency $\varepsilon$-filter (TF $\varepsilon$-filter) is an advanced $\varepsilon$-filter applied to complex spectra along the time axis. It can reduce most kinds of noise while preserving a signal that varies frequently such as a speech signal. The filter design is simple and it can effectively reduce noise. It is applicable not only to small amplitude stationary noise but also to large amplitude nonstationary noise. However when we consider the noise that varies much frequently along the time axis, TF $\varepsilon$-filter cannot reduce noise without the signal distortion. When we consider the noise where the neighboring frequency bins have similar powers such as impulse noise, we can reduce the noise by using $\varepsilon$-filter applied to the complex spectra not along the time axis, but along the frequency axis. This paper introduces an advanced method for noise reduction that applies $\varepsilon$-filter to complex spectra not only along the time axis but also along the frequency axis labeled cross TF $\varepsilon$-filter. We conducted the experiments utilizing the sounds with stationary, nonstationary and natural noise.

1 INTRODUCTION

Noise reduction plays an important role in speech recognition and individual identification. When we consider the instruments like hearing-aids and phones, noise reduction for a monaural sound is strongly expected. It will also be easy to miniaturize the system size because it requires only one signal. The spectral subtraction (SS) is a well-known approach for reducing the noise signal of the monaural-sound (Boll, 1979). It can reduce the noise effectively despite of the simple procedure. However, it can handle only the stationary noise. It also needs to estimate the noise in advance. Although noise reduction utilizing Kalman filter has also been reported (Kalman, 1960; Fujimoto and Ariki, 2002), the calculation cost is large. Some authors have reported a model based approach for noise reduction (Daniel et al., 2006). In this approach, we can extract the objective sound by learning the sound model in advance. However, it is not applicable to the signals with the unknown noise as well as SS. There are some approaches utilizing comb filter (Lim et al., 1978). In this approach, we firstly estimate the pitch of the speech signal, and reduce the noise signal utilizing comb filter. However, the estimation error results in the degradation of the speech quality. Some authors have reported the method utilizing $\varepsilon$-filter (Harashima et al., 1982; Arakawa et al., 2002). $\varepsilon$-filter is a nonlinear filter, which can reduce the noise signal with preserving the signal. $\varepsilon$-filter is simple and has some desirable features for noise reduction. It does not need to have the model not only of the signal but also of the noise in advance. It is easy to be designed and the calculation cost is small. It can reduce not only the stationary noise but also the nonstationary noise. However, it can reduce only the small amplitude noise in principle. To solve the problems, the method labeled TF $\varepsilon$-filter was proposed (Abe et al., 2007). TF $\varepsilon$-filter is an improved $\varepsilon$-filter applied to the complex spectra along the time axis in time-frequency domain. By utilizing TF $\varepsilon$-filter, we can reduce not only small amplitude stationary noise but also large amplitude nonstationary noise. However TF $\varepsilon$-filter cannot reduce the noise without distortion when the noise changes frequently along the time axis such as impulse noise. To solve the problem, we apply $\varepsilon$-filter to complex spectra not only along the time axis but also along the
frequency axis labeled cross TF $\varepsilon$-filter. By applying $\varepsilon$-filter to the complex spectra along the two axes, we can reduce the noise even if it changes frequently along the time axis. It need not estimate the noise as well as TF $\varepsilon$-filter in advance. We also show the experimental results of the proposed method compared to the other methods such as SS and TF $\varepsilon$-filter.

2 NOISE REDUCTION UTILIZING TD $\varepsilon$-FILTER AND TF $\varepsilon$-FILTER

To clarify the problems of a time-domain $\varepsilon$-filter (TD $\varepsilon$-filter)(Harashima et al., 1982; Arakawa et al., 2002), we firstly explain the TD $\varepsilon$-filter algorithm. Let us define $x(k)$ as the input signal at time $k$. Let us also define $y(k)$ as the output signal of the $\varepsilon$-filter at time $k$ as follows:

$$y(k) = x(k) + \sum_{i=-P}^{P} a(i) F(x(k+i) - x(k)), \quad (1)$$

where $a(i)$ represents the filter coefficient. $a(i)$ is usually constrained as follows:

$$\sum_{i=-P}^{P} a(i) = 1. \quad (2)$$

The window size of the $\varepsilon$-filter is $2P + 1$. $F(x)$ is the nonlinear function described as follows:

$$|F(x)| \leq \varepsilon_0 : -\infty < x \leq \infty, \quad (3)$$

where $\varepsilon_0$ is a constant. This method can reduce small amplitude noise while preserving the speech signal. For example, we can set the nonlinear function $F(x)$ as follows:

$$F(x) = \begin{cases} x & (-\varepsilon_0 \leq x \leq \varepsilon_0) \\ 0 & \text{(otherwise)} \end{cases} \quad (4)$$

Figure 1 shows the basic concept of an $\varepsilon$-filter when Eq.4 is utilized as $F(x)$. Figure 1(a) shows the waveform of the input signal.

Executing the $\varepsilon$-filter at point $A$ in Figure 1(a), we first replace all the points where the distance from $A$ is larger than $\varepsilon_0$ by the value of point $A$. We then summate the signals in the same window. Figure 1(b) shows the basic concept of this procedure. The dotted line represents the points where the distance from $A$ is larger than $\varepsilon_0$. In Figure 1(b), the continuous line represents the values replaced through this procedure. As a result, if the points are far from $A$, the points are ignored. On the other hands, if the points are close to $A$, the points are smoothed. Due to this procedure, the $\varepsilon$-filter reduces noise while preserving the precipitous attack and decay of the speech signal. In the same way, by executing the $\varepsilon$-filter at point $B$ in Figure 1(a), we replace all the points where the distance from $B$ is larger than $\varepsilon_0$ by the value of the point $B$. The points are ignored if they are far from $B$, while the points are smoothed if the points are close to $B$ as shown in Figure 1(c). Consequently, we can reduce small amplitude noise near by the processed point while preserving the speech signal.

An $\varepsilon$-filter can reduce small amplitude noise in the time domain. However, due to the procedure, it is not applicable to large amplitude noise. To solve this problem, Time-Frequency $\varepsilon$-filter (TF $\varepsilon$-filter) was proposed (Abe et al., 2007). TF $\varepsilon$-filter utilizes the distribution difference of the speech signal and the noise in the frequency domain. The following assumptions regarding the sound sources are used:

- **Assumption 1.** Speech signal has greater variation in power than noise signal in the time-frequency domain.
- **Assumption 2.** Noise signal is distributed more uniformly and with less variation in the time-frequency domain.

Figure 2 depicts the speech signal and the white noise signal in the time and the time-frequency domains.

As shown in Figure 2, assumptions 1 and 2 are fulfilled in the case of various noise like white noise and natural noise such as the sound of a cooling fan. In Figures 2(b) and (d), the power is normalized using the maximal power of the speech signal. When we consider frequency bins where there are signals, the ratio of noise power to signal power is smaller than the ratio of noise amplitude to signal amplitude in the time domain. In TF $\varepsilon$-filter, we utilize this feature to
Fourier transformation (STFT) as follows:

\[ X(\kappa, \omega) = \sum_{l=-\infty}^{\infty} x(l)W(l)e^{-j\omega l} \]

where \( W \), \( \kappa \), and \( \omega \) represent the window function, the time frame in the time-frequency domain, and the angular frequency, respectively. \( j \) represents the imaginary unit. Next we execute a TF \( \varepsilon \)-filter, which is an \( \varepsilon \)-filter applying to complex spectra along the time axis in the time-frequency domain, as shown in Figure 3(2). In this procedure, \( V(\kappa, \omega) \) is obtained as follows:

\[ V(\kappa, \omega) = \sum_{i=-Q}^{Q} a(i)X'(\kappa + i, \omega), \]

apply an \( \varepsilon \)-filter to high-level noise.

Figure 3 illustrates the method combining TF \( \varepsilon \)-filter and TD \( \varepsilon \)-filter. As shown in Figure 3(1), we firstly transform the input signal \( x(k) \) to the complex amplitude \( X(\kappa, \omega) \) by short term Fourier transformation (STFT) as follows:

\[ X(\kappa, \omega) = \sum_{l=-\infty}^{\infty} x(l)W(l)e^{-j\omega l}, \]

where the window size of \( \varepsilon \)-filter is \( 2Q + 1 \),

\[ X'(\kappa + i, \omega) = \begin{cases} \frac{X(\kappa, \omega)}{\omega} & \frac{|X(\kappa, \omega)| - |X(\kappa + i, \omega)|}{\varepsilon} > \varepsilon_T \\ X(\kappa + i, \omega) & \frac{|X(\kappa, \omega)| - |X(\kappa + i, \omega)|}{\varepsilon} \leq \varepsilon_T \end{cases} \]

and \( \varepsilon_T \) is a constant.

Figure 4 illustrates the differences in performance when we apply a TF \( \varepsilon \)-filter to the speech signal and the noise. The horizontal axis and the vertical axis represent the real axis and the imaginary axis, respectively. In Figure 4, \( * \) and \( \times \) represent the processed point and the other signal points in the same window, respectively. Point A in Figure 4(a) and point B in Figure 4(b) represent the complex amplitude of the processed point. \( A' \) and \( B' \) represent the complex amplitudes of the outputs when we apply the TF \( \varepsilon \)-filter to the points A and B, respectively. Executing the TF \( \varepsilon \)-filter, we firstly replace the complex amplitude of the noise outside of the shadow area by that of A. We then summate the complex spectra of all the points in the same window. Due to handling complex spectra, when we have many signals that have similar amplitudes but different phases, the real part and imaginary part cancel each other. In other words, even if the amplitude of the noise is large, the noise is reduced because they cancel each other. Note that the noise is reduced not only when the amplitude of
the noise is small but also when the amplitude of the noise is large because of this procedure. Figure 4(a) represents the basic concept in the case that the power varies drastically like a speech signal. When we consider a signal whose power varies frequently, the difference between the absolute value of A and that of the other signals is large as shown in Figure 4(a). For this reason, many signals in the same window as the point A are replaced by A. As a result, when we handle the speech signal, the complex amplitude of the processed point is intact. Figure 4(b) represents the basic concept in case that the power does not vary so much like a noise signal. When we consider a noise signal, the difference between the absolute value of B and that of the other signals is relatively small compared with the speech signal. Hence, few signals in the same window as point B are replaced by B. In other words, when handling noise, the complex amplitude of the processed point becomes smaller when the TF ε-filter is applied. Based on these aspects, we can reduce noise while preserving the signal by setting εT appropriately.

Hence, the TF ε-filter is effective even when the power of the noise to signal is large. Additionally, under assumption 2, the TF ε-filter becomes more effective. When assumption 2 is satisfied, the variation of the noise to the signal in the frequency domain becomes smaller than that in the time domain. As a consequence, even if the noise varies frequently in the time domain, the ε-filter can be applied in the time-frequency domain.

Next, we transform $V(\kappa, \omega)$ to $v(k)$ by inverse STFT as shown in Figure 3(3).

To reduce the remaining noise, we additionally apply the ε-filter in the time domain to $v(k)$ as shown in Figure 3(4). Note that the ε-filter in the time domain can be utilized because large amplitude noise has already been reduced in the previous procedure. The output $y(k)$ can be obtained as follows:

$$y(k) = \sum_{i=-P}^{P} a(i) v'(k+i),$$

where

$$v'(k+i) = \begin{cases} v(k) & (|v(k+i) - v(k)| > \varepsilon_t) \\ v(k+i) & (|v(k+i) - v(k)| \leq \varepsilon_t) \end{cases}$$

and $\varepsilon_t$ is a constant.

### 3 NOISE REDUCTION UTILIZING CROSS TF ε-FILTER

TF ε-filter can reduce the various types of noise effectively. However, when we use the noise that varies much frequently along the time axis, TF ε-filter cannot reduce noise without the signal distortion. When we consider the noise where the neighboring frequency bins have similar powers such as impulse noise, we can reduce the noise by using ε-filter applying to complex spectra not along the time axis, but along the frequency axis.

Figure 5 shows the basic concept of the proposed method. At first, as shown in Figure 5 “Step 1”, we apply the ε-filter to the complex spectra along the frequency axis to reduce the noise where the neighboring frequency bins have similar power although the noise amplitude varies drastically such as the impulse noise and white noise whose amplitude varies. Next we apply ε-filter to complex spectra along the time axis to reduce the noise that distributes wider than the speech signal in the frequency domain as shown in Figure 5 “Step 2”.

Figure 6 illustrates the proposed method with a
Figure 6: The block diagram of the proposed method.

block diagram. Let us consider $x(k)$, and $X(\kappa, \omega)$
transformed from $x(k)$ by STFT as well as Eq.5 as shown in Figure 6(1). Next we apply $\varepsilon$-filter to com-
plex spectra along the frequency axis as shown in Fig-
ure 6(2). In this procedure, $U(\kappa, \omega)$ is obtained as follows:

$$U(\kappa, \omega) = \sum_{i=-N}^{N} a(i) X'(\kappa + i, \omega), \quad (10)$$

where

$$X'(\kappa + i, \omega)
= \begin{cases} X(\kappa, \omega) & \text{if} \ |X(\kappa + i, \omega)| - |X(\kappa, \omega + i)| > \varepsilon_F \\ X(\kappa, \omega + i) & \text{if} \ |X(\kappa + i, \omega)| - |X(\kappa, \omega + i)| \leq \varepsilon_F \end{cases}. \quad (11)$$

$\varepsilon_F$ is a constant and $2N + 1$ is window size. Then we employ $\varepsilon$-filter applying to complex spectra along the
time axis as shown in Figure 6(3). In this procedure, $Y(\kappa, \omega)$ is obtained as follows:

$$Y(\kappa, \omega) = \sum_{i=-M}^{M} a(i) U'(\kappa + i, \omega), \quad (12)$$

where

$$U'(\kappa + i, \omega)
= \begin{cases} U(\kappa, \omega) & \text{if} \ |U(\kappa + i, \omega)| - |U(\kappa + i, \omega + i)| > \varepsilon_T \\ U(\kappa + i, \omega + i) & \text{if} \ |U(\kappa + i, \omega)| - |U(\kappa + i, \omega + i)| \leq \varepsilon_T \end{cases}. \quad (13)$$

$\varepsilon_T$ is a constant and $2M + 1$ is the window size. Next we transform $Y(\kappa, \omega)$ to $y(k)$ by inverse STFT as shown in Figure 6(4). We label this process “cross
TF $\varepsilon$-filter”.

4 EXPERIMENT

4.1 Experimental Condition

We conducted the experiments utilizing a speech signal
with a noise signal. As the speech signal, we uti-

\begin{table}[h]
\centering
\caption{Common parameters.}
\label{tab:parameters}
\begin{tabular}{|l|c|}
\hline
Parameter & Value \\
\hline
Sampling frequency & 44100 \\
Block size & 512 \\
Hop size & 256 \\
Window function & Hanning window \\
\hline
\end{tabular}
\end{table}

lized “Japanese Newspaper Article Sentences” edited
by the Acoustical Society of Japan. We also prepared
three kinds of noise signals: stationary noise, nonsta-
tionary noise and natural noise. The signal and the
noise are mixed in the computer. To compare the ef-
fectiveness of the proposed method to other methods,
we conducted the experiments utilizing three meth-
ods; spectral subtraction (SS), the method combining
TF $\varepsilon$-filter and TD $\varepsilon$-filter and cross TF $\varepsilon$-filter. Table
1 shows the value of common parameters for all the
experiments.

To evaluate the performance of noise reduction,
we use signal-to-noise ratio ($SNR$) and signal-to-
distortion ratio ($SDR$). $SNR$ is defined as follows:

$$SNR = 10 \cdot \log_{10} \left( \frac{\sum_{k=1}^{L} s(k)^2}{\sum_{k=1}^{L} n(k)^2} \right), \quad (14)$$

where $s(k)$, $n(k)$ and $L$ represent the speech signal at
time $k$, the noise signal at time $k$, and the length of the
signal, respectively. To calculate $SDR$ of the output
signal, we separately applied each method to the signal
and noise, and calculated the output $SNR$ by using
the obtained signal and noise. $SDR$ can be represented
as follows:

$$SDR = 10 \cdot \log_{10} \left( \frac{\sum_{k=1}^{L} s_{\text{in}}(k)^2}{\sum_{k=1}^{L} (s_{\text{in}}(k) - s_{\text{out}}(k))^2} \right), \quad (15)$$

where $s_{\text{in}}(k)$ and $s_{\text{out}}(k)$ represent the input signal
and the output signal at time $k$ respectively, when we used
only the speech signal. $SDR$ represents how much the
signal is distorted by reducing the noise. Throughout
all the experiments, the parameters of SS and the
method combining TF $\varepsilon$-filter and TD $\varepsilon$-filter are set
optimally. On the other hand, in cross TF $\varepsilon$-filter, $\varepsilon_T$
was set at 0.1. We only change $\varepsilon_T$ depending on the
noise to show the robustness concerning the parameter
setting although we can reduce the noise more
effectively. $SNR$ of the input signal is set at 10[dB]
throughout all of the experiments.
Table 2: SNR and SDR when a signal with stationary noise is utilized.

<table>
<thead>
<tr>
<th></th>
<th>SNR[dB]</th>
<th>SDR[dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input signal</td>
<td>10.0</td>
<td>—</td>
</tr>
<tr>
<td>SS</td>
<td>18.3</td>
<td>21.7</td>
</tr>
<tr>
<td>TD ε-filter processed after TF ε-filter</td>
<td>40.6</td>
<td>18.9</td>
</tr>
<tr>
<td>Cross TF ε-filter</td>
<td>44.4</td>
<td>19.3</td>
</tr>
</tbody>
</table>

4.2 Experimental Results in the Case of Stationary Noise

We first conducted the experiment utilizing a signal with stationary noise. We prepared a speech signal and white noise as the signal and the stationary noise, respectively. We set \( \varepsilon_F \) in the proposed method at 0.7. We also set the window size of ε-filter applied to complex spectra in the proposed method along the frequency axis and the time axis at 101 and 11, respectively. Table 2 shows the results of the experiments for stationary noise.

As shown in Table 2, the proposed method could reduce the noise compared to the other methods with preserving the signal. Figure 7 shows the sound spectrograms. In Figure 7, bright color represents that the signal power is high while dark color represents that the signal power is low. Figure 7(a) shows the spectrogram of the original signal. Figure 7(b) shows the spectrogram of the signal with stationary noise. Figures 7(c)-(e) show the spectrograms of the output of SS, the output of the method combining TF ε-filter and TD ε-filter and the output of the proposed method, respectively. As shown in Figure 7, when we used the proposed method, the noise could be reduced more effectively than the other methods.

4.3 Experimental Results in the Case of Nonstationary Noise

The experiment was conducted using a signal with nonstationary noise. We used the same speech signal as in Sec.4.2. We prepared white noise with an amplitude that sometimes varied. We set \( \varepsilon_F \) in the proposed method at 1.1. We also set the window size of ε-filter applied to complex spectra in the proposed method along the frequency axis and the time axis 81 and 11, respectively. Table 3 shows the results of the experiments on nonstationary noise. As shown in Table 3, the SNR of the proposed method is superior to those of the other methods. Figure 8(a) shows the spectrogram of the original signal. Figure 8(b) shows the spectrogram of the signal with nonstationary noise. Figures 8(c)-(e) show the spectrograms of the outputs of SS, the output of the method combining TF ε-filter and TD ε-filter and the output of the proposed method, respectively. The relation between the color and signal power is the same as in Sec.4.2. As shown in Figure 8, when we used the proposed method, the noise could be reduced more effectively than the other methods even if we use the nonstationary noise as noise.

4.4 Experimental Results in the Case of Natural Noise

To evaluate the performance of the proposed method for natural noise, we conducted the experiment utilizing a speech signal and a noise generated from the cooling fan of a personal computer. The most power of noise used in this experiment is distributed in the low-frequency range. We set \( \varepsilon_F \) in the proposed method at 1.8. We also set the window size of ε-filter applied to complex spectra in the proposed method along the frequency axis and the time axis 51 and 11.
Table 4: SNR and SDR when a signal with natural noise is utilized.

<table>
<thead>
<tr>
<th></th>
<th>SNR[dB]</th>
<th>SDR[dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input signal</td>
<td>10.0</td>
<td>—</td>
</tr>
<tr>
<td>SS</td>
<td>17.4</td>
<td>20.1</td>
</tr>
<tr>
<td>TD ε-filter processed after TF ε-filter</td>
<td>38.2</td>
<td>13.4</td>
</tr>
<tr>
<td>Cross TF ε-filter</td>
<td>40.2</td>
<td>15.0</td>
</tr>
</tbody>
</table>

respectively. Table 4 shows the results of the experiments for natural noise. As shown in Table 4, the SNR of the proposed method is superior to those of the other methods as well as in the case of stationary noise and nonstationary noise. Figure 9(a) shows the spectrogram of the original signal. Figure 9(b) shows the spectrogram of the signal with natural noise. Figures 9(c)-(e) show the spectrograms of the outputs of SS, the output of the method combining TF ε-filter and TD ε-filter and the output of the proposed method, respectively. The relation between the color and signal power is the same as in Sec.4.2. As shown in Figure 9, when we use the proposed method, the noise could be reduced more effectively than the other methods even if the natural noise was used as noise.

5 CONCLUSIONS

In this paper, we introduced an algorithm for noise reduction applying ε-filter to complex spectra not only along the time axis but also along the frequency axis in time-frequency domain. The proposed method can reduce not only stationary noise but also nonstationary and natural noise effectively with preserving signal clarity. The experimental results showed that the proposed method could be applied to various kinds of noise. The proposed method could reduce the louder noise compared with the conventional methods such as the method combining TF and TD ε-filter and SS. It is considered that the proposed method can be applied not only to the speech in Japanese but also to the speech in English with noise because the performance of the proposed method depends on only the power change of speech and noise signal. For future works, we would like to confirm the robustness of the proposed method for various input SNR and various types of noise. We also aim to determine each parameter adaptively.

ACKNOWLEDGEMENTS

This research was supported by the research grant of Support Center for Advanced Telecommunications Technology Research (SCAT), by the research grant of Tateisi Science and Technology Foundation, and by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 20700168, 2008. This research was also supported by Waseda University Grant for Special Research Projects(B), No.2007B-142, No.2007B-143 and No.2007B-168, by "Establishment of Consolidated Research Institute for Advanced Science and Medical Care", Encouraging Development Strategic Research Centers Program, the Special Coordina-
tion Funds for Promoting Science and Technology, Ministry of Education, Culture, Sports, Science and Technology, Japan, by the CREST project "Foundation of technology supporting the creation of digital media contents" of JST, by the Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture, the 21st Century Center of Excellence Program, "The innovative research on symbiosis technologies for human and robots in the elderly dominated society", Waseda University and by Project for Strategic Development of Advanced Robotics Elemental Technologies (NEDO: 06002090).

REFERENCES


