Simulated Annealing based Parameter Optimization of Time-frequency ε-filter Utilizing Correlation Coefficient

Tomomi Matsumoto[1], Mitsuharu Matsumoto[2] and Shuji Hashimoto[1]

[1]*Department of Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo, Japan*

[2]*The Education and Research Center for Frontier Science, University of Electro-communications, 1-5-1, Chofugaoka, Chofu-shi, Tokyo, Japan*

Keywords: Simulated Annealing, Parameter Optimization, Noise Reduction, ε-filter, Nonlinear Filter, Time-frequency ε-filter.

Abstract: Time-Frequency ε-filter (TF ε-filter) can reduce different types of noise from a single-channel noisy signal while preserving the signal that varies drastically such as a speech signal. It can reduce not only small stationary noise but also large nonstationary noise. However, it has some parameters whose values are set empirically. So far, there are few studies to optimize the parameter of TF ε-filter automatically. In this paper, we employ the correlation coefficient of the filter output and the difference between the filter input and output as the evaluation function of the parameter optimization. We also propose an algorithm to set the optimal parameter of TF ε-filter automatically. The experimental results show that we can obtain the adequate parameter in TF ε-filter automatically by using the proposed method.

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 single-channel signal is required. The spectral subtraction (SS) is a well-known approach for reducing the noise signal of the monaural-sound (Boll, 1979; Lim, 1978). It can reduce the noise effectively with the simple procedure. However, it can handle only the stationary noise. It also needs to estimate the noise spectrum 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 constructing the sound model in advance. However, it is not applicable to the signals with the unknown noise. 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 especially in case of consonant.

Harashima et al. have reported a nonlinear filter named ε-filter , which can reduce noise while preserving the signal (Harashima et al., 1982) . We label it "TD ε-filter" as it handles signal shape in time domain. TD ε-filter is simple and has some desirable features for noise reduction. It does not require 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 time-frequency ε-filter (TF ε-filter) was proposed (Abe et al., 2007). TF ε-filter is an improved ε-filter applied to the complex spectra along the time axis in time-frequency domain. By utilizing TF ε-filter, we can reduce not only small amplitude stationary noise but also large amplitude nonstationary noise. However, TF ε-filter has some parameters and we need to set them adequately based on empirical control. Moreover, as we only have a single-channel noisy signal, it is difficult to evaluate whether the parameter is optimal or not. We cannot know the difference between the original signal and the filter output from the observed signal. So far, there are few studies on the appropriateness of the parameter setting of TF ε-filter.

Based on the above prospects, we proposed an ap-

Matsumoto T., Matsumoto M. and Hashimoto S..
Simulated Annealing based Parameter Optimization of Time-frequency e-filter Utilizing Correlation Coefficient.
DOI: 10.5220/0004126602370241
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications and Wireless Information Networks and Systems* (SIGMAP-2012), pages 237-241
ISBN: 978-989-8565-25-9

proach to set the parameter (Abe et al., 2009). In this method, as a simple criterion, we assume that the signal and noise are noncorrelated. And we employ the correlation coefficient of the filter output and the difference between the input signal and the filter output to set the parameter adequately. In this study, we confirmed that the correlation became almost minimal when the error was minimal. However, as the correlation has many local minimum, we could not employ gradient method. To solve the problem, we use simulated annealing to set the parameter in this paper. When we utilize the proposed method, we can set the parameter adequately without the information about the noise and the signal. In Sec.2, we explain TF ε-filter to clarify the problem. In Sec.3, we describe the algorithm of the method to determine the parameter adequately. In Sec.4, we show the experimental results. Experimental results show that the proposed method can estimate the optimal parameter of the TF ε-filter. Conclusions are given in Sec.5.

2 TIME-FREQUENCY ε-FILTER

In this section, we briefly describe the TF ε-filter algorithm. TF ε-filter is an improved ε-filter applied to the complex spectra along the time axis in time-frequency domain.

Let us define x_k as the input signal sampled at time k. In TF ε-filter, we firstly transform the input signal x_k to the complex spectra $X_{\kappa,\omega}$ by short term Fourier transformation (STFT). κ and ω represent the time frame and the angular frequency in the time-frequency domain, respectively. κ and ω are integer numbers. Next we execute a TF ε-filter, which is an ε-filter applying to complex spectra along the time axis in the time-frequency domain. In this procedure, $Y_{\kappa,\omega}$ is obtained as follows:

$$Y_{\kappa,\omega} = \sum_{i=-Q}^{Q} \frac{1}{2Q+1} X'_{\kappa+i,\omega}, \tag{1}$$

where

$$X'_{\kappa+i,\omega} = \begin{cases} X_{\kappa,\omega} & (||X_{\kappa,\omega}| - |X_{\kappa+i,\omega}|| > \varepsilon) \\ X_{\kappa+i,\omega} & (||X_{\kappa,\omega}| - |X_{\kappa+i,\omega}|| \le \varepsilon), \end{cases} \tag{2}$$

and ε is a constant. We define the window size of ε-filter as $2Q+1$. Then, we transform $Y_{\kappa,\omega}$ to the output signal y_k by inverse STFT.

By utilizing TF ε-filter, we can reduce not only small amplitude stationary noise but also large amplitude nonstationary noise because the noise power is distributed in wide frequency range even when the noise has large amplitude in time domain. It does not require either the model of the signal or that of the noise in advance. It is easy to be designed and the calculation cost is small (Abe et al., 2007).

3 AUTOMATIC PARAMETER OPTIMIZATION UTILIZING CORRELATION COEFFICIENT

As described in the previous section, when the TF ε-filter is employed, we need to set ε value adequately to reduce the noise. However, we cannot estimate the optimal parameter because the noise and signal are not known throughout all the procedures.

To solve the problem, we pay attention to the correlation of the target signal and the noise signal. We make the following assumption concerning the target signal and noise signal:

- **Assumption 1.** The target signal is noncorrelated with the noise signal.

Let us define s_k and n_k as the objective signal and the noise signal, respectively. Let $R(s_k, n_k)$ be the correlation coefficient of s_k and n_k described as follows:

$$R(s_k, n_k) = \frac{\sum_{k=1}^{L} (s_k - \overline{s_k})(n_k - \overline{n_k})}{\sqrt{\sum_{k=1}^{L} (s_k - \overline{s_k})^2} \sqrt{\sum_{k=1}^{L} (n_k - \overline{n_k})^2}}, \tag{3}$$

where L is the data length. $\overline{s_k}$ and $\overline{n_k}$ represent the averages of s_k and n_k, respectively. $\overline{s_k}$ and $\overline{n_k}$ are described as follows:

$$\overline{s_k} = \frac{1}{L} \sum_{k=1}^{L} s_k. \tag{4}$$

$$\overline{n_k} = \frac{1}{L} \sum_{k=1}^{L} n_k. \tag{5}$$

When L is large enough, it is expected that the following equation is satisfied under assumption 1:

$$R(s_k, n_k) = 0. \tag{6}$$

As described above, s_k and n_k are unknown throughout the filtering procedures. Instead of s_k and n_k, we consider the correlation coefficient of the filter output and the difference between the input signal and the filter output. Let us consider x_k and y_k as the input signal and the output signal of TF ε-filter, respectively. x_k can be described as follows:

$$x_k = s_k + n_k. \tag{7}$$

When the TF ε-filter can reduce the whole noise, while it preserves the signal completely, the filter output y_k equals the signal s_k. The noise n_k can be described as follows:

$$\begin{aligned} n_k &= x_k - s_k \\ &= x_k - y_k. \end{aligned} \tag{8}$$

Although actual TF ε-filter does not reduce the whole noise and reduces the signal, if ε value is set optimally, it is expected that the correlation of y_k and $x_k - y_k$ becomes smaller than that of y_k and $x_k - y_k$ in other ε. Hence, the optimal parameter ε_{opt} can be obtained as

$$\varepsilon_{opt} = \arg\min_{\varepsilon} |R(y_k, x_k - y_k)|, \tag{9}$$

where

$$R(y_k, x_k - y_k) \tag{10}$$
$$= \frac{\sum_{k=1}^{L}(y_k - \overline{y_k})(x_k - y_k - \overline{x_k - y_k})}{\sqrt{\sum_{k=1}^{L}(y_k - \overline{y_k})^2}\sqrt{\sum_{k=1}^{L}(x_k - y_k - \overline{x_k - y_k})^2}},$$

where $\overline{x_k}$ and $\overline{x_k - y_k}$ represent the average of x_k and $x_k - y_k$, respectively. $\overline{x_k}$ and $\overline{x_k - y_k}$ are described as follows:

$$\overline{x_k} = \frac{1}{L}\sum_{k=1}^{L} x_k. \tag{11}$$

$$\overline{x_k - y_k} = \frac{1}{L}\sum_{k=1}^{L}(x_k - y_k). \tag{12}$$

To obtain the adequate parameter automatically, we utilize the simulated annealing. The process of the simulated annealing is represented as follows:

Step1 We set ε at the initial parameter ε_0 and the initial temperature T.

Step2 We calculate the initial solution $y_k(\varepsilon_0)$.

Step3 We repeat the following process until the termination condition is fulfilled.

1. We randomly choose the ε′ which satisfies: $\varepsilon - a < \varepsilon' < \varepsilon + a$. Where a is the constant number which constrains ε′ within the neighborhood of ε.
2. We calculate the $|R(y_k(\varepsilon), x_k - y_k(\varepsilon))|$ and $|R(y_k(\varepsilon'), x_k - y_k(\varepsilon'))|$
3. If $|R(y_k(\varepsilon'), x_k - y_k(\varepsilon'))| \leq |R(y_k(\varepsilon), x_k - y_k(\varepsilon))|$, ε is replaced to ε′. Otherwise ε is replaced to ε′ with probability $e^{-(y_k(\varepsilon') - y_k(\varepsilon))/T}$.

Step4 Step 2 and 3 are repeated for a while. If ε value is kept despite the procedure, the terminal condition of iteration is fulfilled. And we regard the ε as the optimized solution. Otherwise we decrease T and go back to the Step2.

As the initial parameter ε_0, we use the value described by the following equation.

$$\varepsilon_0 = \overline{\sigma(X_{\kappa,\omega})} \tag{13}$$

where

$$\overline{\sigma(X_{\kappa,\omega})} = \frac{1}{M}\sum_{\omega=1}^{M}\sqrt{\frac{1}{N}\sum_{\kappa=1}^{N}(X_{\kappa,\omega} - \overline{X_{\kappa,\omega}})^2}, \tag{14}$$

where M and N represent the number of frequency resolution and the number of the time frame, respectively. Eq. 14 represents the mean along the frequency axis of the standard deviation along the time axis of $X_{\kappa,\omega}$. This is because the standard deviation represents the fluctuation of $N_{\kappa,\omega}$ that is the transformed n_k by FFT when x_k is equal to n_k. Therefore, it is considered that most of noise will be reduced by the TF ε-filter under the above situation. In practice, x_k includes s_k and therefore it does not correspond to the correct fluctuation of n_k. However, we consider that the standard deviation is useful as a first order approximation of ε when we only have the input signal x_k and filter output y_k.

4 EXPERIMENT

To clarify the adequateness of the proposed method, we conducted the experiments utilizing monaural sounds with the speech signal and the noise signal. In the experiments, we updated the ε value by using the proposed method and checked whether the proposed method worked well. In the experiments, we calculate $R(y_k, x_k - y_k)$ and the mean square error (MSE) between the original signal s_k and the filter output y_k. MSE is defined as follows:

$$MSE = \frac{1}{L}\sum_{k=1}^{L}(s_k - y_k)^2. \tag{15}$$

As the sound source, we used “Japanese Newspaper Article Sentences” edited by the Acoustical Society of Japan. We used the white noise with uniform distribution as the stationary noise. When ε is too small, the difference between the input and the filter output becomes small. Due to this reason, $R(y_k, x_k - y_k)$ becomes close to 0 when ε is too small. Hence, we constrained ε larger than 0.01.

Figure 1 shows the relation between ε, correlation coefficient and MSE. As shown in Fig.1, MSE became minimal when we set ε to 0.42. At this time, the correlation coefficient also became almost minimal.

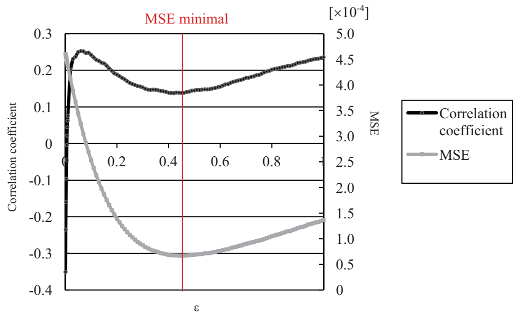
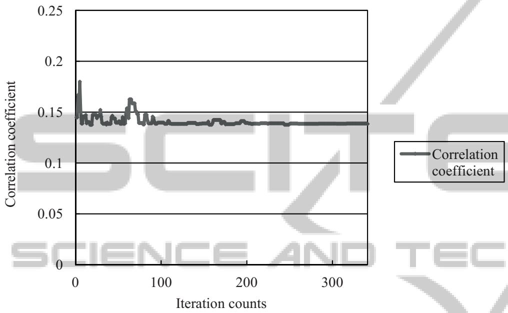
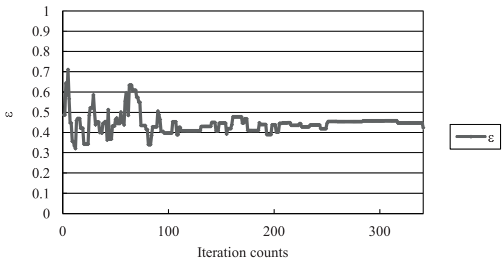


Figure 1: Relation between ε, correlation coefficient and MSE.



(a) Relation between iteration counts and correlation coefficient.

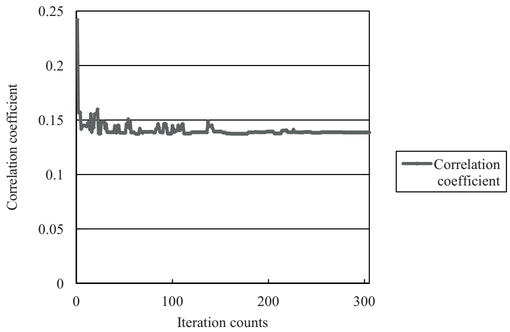


(b) Relation between iteration counts and ε.

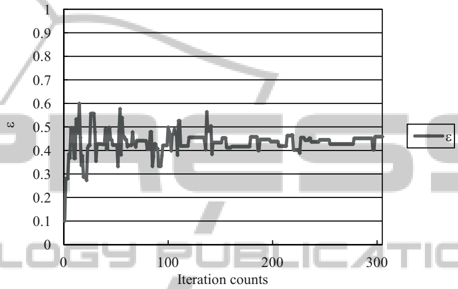
Figure 2: Transition of correlation coefficient and ε when we set the initial ε to 0.49.

Figure 2 shows the transition of correlation coefficient and ε when we set the initial ε to 0.49, that was obtained by Eq.14. As shown in Fig.2, we can obtain the adequate ε utilizing the proposed method automatically.

To show the robustness for changing the initial ε, we conducted the experiments using the different initial ε. Figures 3 and 4 show the transition of correlation coefficient and ε when we set the initial ε to 0.1 and 0.9, respectively. As shown in Figs.3 and 4, we can obtain the adequate parameter of ε utilizing the proposed method even when the initial ε is much larger or smaller than the optimal ε.



(a) Relation between iteration counts and correlation coefficient.

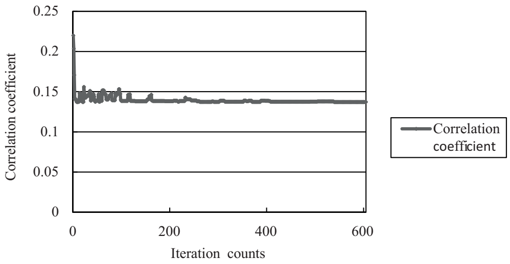


(b) Relation between iteration counts and ε.

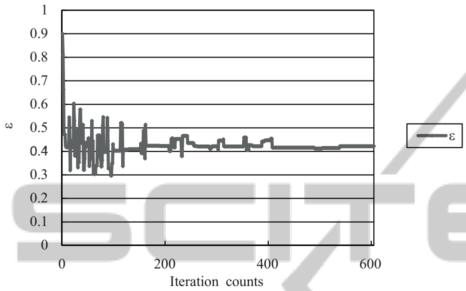
Figure 3: Transition of correlation coefficient and ε when we set the initial ε to 0.1.

5 CONCLUSIONS

In this paper, we employed the correlation coefficient of the filter output and the difference between the input and the filter output as the evaluation function of the parameter setting of TF ε-filter. We also proposed a simulated annealing based algorithm to determine the parameter of TF ε-filter automatically. The experimental results show that we can automatically determine the adequate parameters of TF ε-filter by utilizing our method. As the proposed method only assumes the decorrelation of the signal and noise, it is expected that the application range of the proposed method is large. Although we only have the single-channel noisy signal, our method enables us to obtain an adequate ε parameter automatically. The proposed method does not require an estimation of the noise in advance. The features will help us to use TF ε-filter in a practical situation. To handle nonstationary noise, we need to change ε adaptively depending on the noise. Hence, we aim to improve our method to solve this problem. For future studies, we would like to evaluate robustness when changing the window size of the TF ε-filter. We also would like to determine all parameters in TF ε-filter, that is, not only



(a) Relation between iteration counts and correlation coefficient.



(b) Relation between iteration counts and ε.

Figure 4: Transition of correlation coefficient and ε when we set the initial ε to 0.9.

the ε value but also the window size adequately based on automatic control.

ACKNOWLEDGEMENTS

This research was supported by Special Coordination Funds for Promoting Science and Technology, by Japan Prize Foundation, NS promotion foundation for science of perception and Foundation for the Fusion Of Science and Technology, by Special Coordination Funds for Promoting Science and Technology, and by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 22700186, 2010. This research was also supported by the CREST project "Foundation of technology supporting the creation of digital media contents" of JST, and by the Global-COE Program,"Global Robot Academia", Waseda University.

REFERENCES

Abe, T., Matsumoto, M., and Hashimoto, S. (2007). Noise reduction combining time-domain ε-filter and time-frequency ε-filter. In *J. of the Acoust. Soc. America.*, volume 122, pages 2697–2705.

Abe, T., Matsumoto, M., and Hashimoto, S. (2009). Parameter optimization in time-frequency -filter based on correlation coefficient. In *Proc. of International conference on signal processing and multimedia applications (SIGMAP2009)*, pages 107–111.

Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. In *IEEE Trans. Acoust. Speech Signal Process.*, volume ASSP-27, pages 113–120.

Daniel, P., Ellis, W., and Weiss., R. (2006). Model-based monaural source separation using a vector-quantized phase-vocoder representation. In *Proc. IEEE Int'l Conf. on Acoustics, Speech, and Signal Process. 2006*.

Fujimoto, M. and Ariki, Y. (2002). Speech recognition under noisy environments using speech signal estimation method based on kalman filter. In *IEICE Trans. Information and Systems*, volume J85-D-II, pages 1–11.

Harashima, H., Odajima, K., Shishikui, Y., and Miyakawa, H. (1982). ε-separating nonlinear digital filter and its applications. In *IEICE trans on Fundamentals.*, volume J65-A, pages 297–303.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. In *Trans. of the ASME*, volume 82, pages 35–45.

Lim, J. S. (1978). Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise. In *IEEE Trans. Acoust. Speech Signal Process.*, volume ASSP-26, pages 471–472.

Lim, J. S., Oppenheim, A. V., and Braida, L. D. (1978). Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition. In *IEEE Trans. on Acoust. Speech Signal Process.*, volume ASSP-26, pages 419–423.