The 23rd International Technical Conference on Circuits/Systems,
Computers and Communications (ITC-CSCC 2008)

Adaptive Noise Reduction Filter for Speech Using Cascaded Sandglass-type Neural Network

Hiroki Yoshimura†, Tadaaki Shimizu†, Toshie Matumura††, Masaya Kimoto†, and Naoki Isu†††

† Department of Information and Knowledge Engineering, Tottori University, Japan
†† Information Engineering, Nara National College, Japan
††† Department of Information Engineering,Mie University, Japan
†4-101 Koyama-minami, Tottori, Tottori 680-8552, Japan
††22 Yada, Yamato-kohriyama, Nara 639-1080, Japan
†††1577 Kuria-machiya, Tsu, Mie 514-8507, Japan
† Tel. +81-857-31-5107 Fax: +81-857-31-0879
†† Tel. +81-743-6000 Fax: +81-743-55-6019
††† Tel. +81-59-231-9458 Fax: +81-59-231-9458

E-mail: †yosimura@ike.tottori-u.ac.jp

Keywords: Speech Signal, Neural Network, Adaptive Filter, Noise Reduction

1. Introduction

As the number of elderly people is increasing , the demand for high-quality hearing aids is growing worldwide. Large numbers of studies have improved hearing-aid quality for many years[1]–[3]. The hearing aid is required to have frequency characteristics in accordance with both the auditory perception of the user and the characteristics of the input speech signal. This requirement is implemented by a non-linear amplifying function that is composed of tensile and compression amplification[1]. The noise properties are always changing with the time and place where the hearing aid is used.

A hearing aid can never be good enough even if it reduces noise and enhances the amplitude of the speech signal. Hearing aids are required to output clearly audible speech sound. This performance is achieved by preprocessing using an adaptive noise reduction filter[1]. In order to develop a digital hearing aid, we employed an adaptive noise reduction filter as a sandglass-type neural network (SNN) [4]–[6].

In this study, we developed a new adaptive noise reduction filter (CSNNFR: Cascaded Sandglass-type Neural Network Noise Reduction Filter) which is extended from the SNNRF (Sandglass-type Neural Network Noise Reduction Filter). The CSNNRF can perform adaptive noise reduction while capturing dynamic characteristics of the speech signal, owing to the plasticity of the SNN. The CSNNRF is suitable for application to the hearing aid since the ease of hearing the speech signal after noise reduction is more important than the denoising rate. We examined the improvement rate of SN ratio and performed a hearing

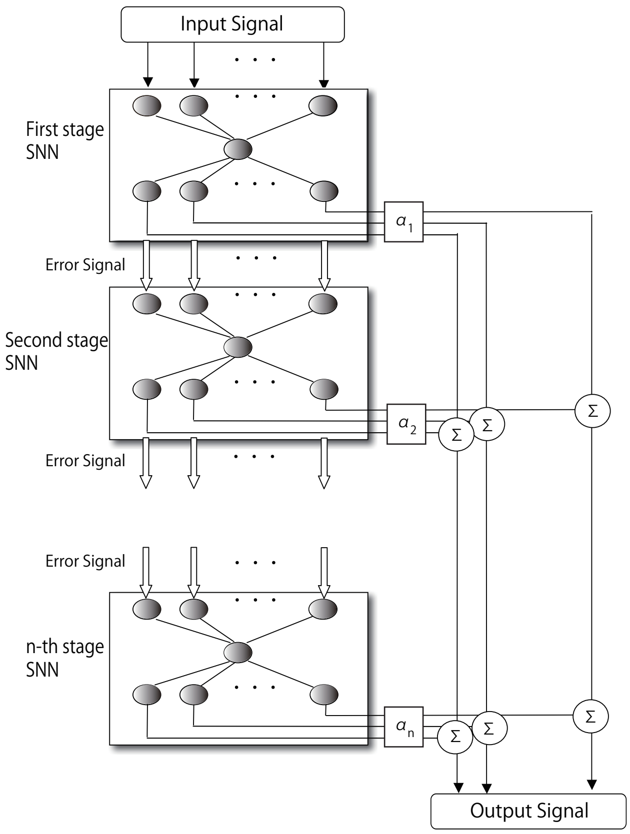

Figure 1 Structure of the CSNNRF

experiment to evaluate the performance of the CSNNRF as a hearing aid.

2. CSNNRF Configuration

We proposed a SNNRF with a structure connecting SNNs in a cascade (CSNNRF) (Fig.1). The SNN is a hierarchy model neural network whose structure has features where by the number of output layer units is the same as that of that of output layer units, and the number of hidden layer units is less than the number of input and output layer units. The response functions of units of the SNN are linear functions. Then, the SNN excutes a transform equivalent to the Karhunen-Loeve (KL) transform. In other words, it is known that the SNN works equivalently to principal component analysis. When the hidden layer of SNN has M units, the highest M components of the KL transform involved in the input signal are extracted as its output signal.

In this study, each SNN has one unit in the hidden layer and 20 units in the input and output layers. The number of cascade connections of SNNs is 20 stages. A speech signal including white noise is put into the first stage SNN of the CSNNRF. In SNNs of second and later stages, error signals generated in the preceding stages (difference in signal between the input and output of the SNN) are given as input signals. Then, the output layer of the n-th stage SNN outputs a time-series signal of the n-th principal component. The CSNNRF divides the input signal into principal components in the order of the rate of contribution. The output signal of the CSNNRF is given as a sum after multiplying the output signal of each i-th SNN by α_i, the attenuation coefficient, which is described in the next section.

3. The filtering of the speech signal by the CSSNRF

3.1 A power estimate using the standard model group of the vowel

In Japanese, the time ratio of a voiced sound such as a vowel in a speech signal is high, and there is a large difference in the power of each principal component (the contribution rate). On the other hand, the white noise is equal in all contribution rates of each principal component. In the CSNNRF, the contribution rate of each principal component of the observation signal is provided from the output of the hidden unit of the SNN of each stage. Using the differences in the characteristics of the contribution rate of speech signals and the white noise, a power ratio of speech and noise signal can be estimated. Accordingly, the standard model group of the vowel was previously generated by principal component analysis of consecutive speech signals without noise involving 24 speakers in total (12 men, 12 women).

We assume the standard model of the speech component $x_i^{(m)}$ (i stands for i-th stage SNN, m stands for m-th model), and an estimate of the power of a speech and noise component for the model, $\hat{S}^{(m)}$ and $\hat{N}^{(m)}$, respectively. Then, the estimate of the power of the principal component $\hat{y}_i^{(m)}$ is as follows:

$$\hat{y}_i^{(m)} = \hat{S}^{(m)} x_i^{(m)} + \hat{N}^{(m)} \tag{1}$$

A sum of the square error of y_i from its $\hat{y}_i^{(m)}$ is obtained for every m-th model as follows:

$$\varepsilon^{(m)} = \sum_i \left(y_i - \hat{y}_i^{(m)} \right)^2 \tag{2}$$

$\hat{S}$ and $\hat{N}$ are given by $\hat{S}^{(m)}$ and $\hat{N}^{(m)}$ of the m-th model whose $\varepsilon^{(m)}$ is the smallest.

3.2 Decision regarding attenuation coefficient α_i

From the expectation that the contribution of all principal components for white noise is equal, when the stage number is R in the CSNNRF, the power of the noise components of each principal components is estimated as follows:

$$\hat{N}_i = \hat{N} / R \tag{3}$$

From $\hat{N}_i$ and y_i, the power of the signal component $\hat{S}_i$ is calculated as follows:

$$\begin{aligned} \hat{S}_i &= E\left[S_i \mid y_i, \hat{N}_i \right] \\ &= \alpha_i y_i \end{aligned} \tag{4}$$

$$\alpha_i = \frac{\hat{S}_i}{\hat{S}_i + \hat{N}_i} = \frac{y_i - \hat{N}_i}{y_i} \tag{5}$$

where α_i is the attenuation coefficient of the i-th stage SNN. The output signal of the CSNNRF is a summation after multiplying the output signal of each i-th SNN by α_i. If the ratio of noise components included is large, the value of α_i is small, and if the ratio of speech components is large, the value of α_i is large. As a result, the CSNNRF can reduce the noise component.

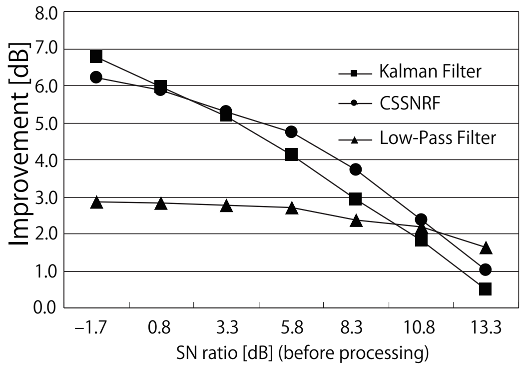

Figure2 SN ratio improvement

4. Performance evaluation of the filtering processing

We clarified the utility of the CSNNRF by examining the SN ratio improvement rate and performance in a hearing experiment.

A speech sample including noise was input into the CSNNRF, and the SN ratio improvement rate on filter processing was obtained from the output signal. For performance comparison, a Kalman filter and a low-pass filter were compared. The cutoff frequancy for a low-pass filter is 4000[Hz]. Figure 2 shows the results of SN ratio improvement. In the case of a range from 3.3 [dB] to 10.8 [dB] in the SN ratio of the input signal, the CSNNRF achieved a higher SN ratio improvement than Kalman filter.

We performed a hearing experiment to investigate the following in order to evaluate CSNNRF performance as a hearing aid from a number of perspectives:

a. "Comfortable sound"
b. "Clarity"
c. "Raucousness of the noise"

Scheffe's method of paired comparisons was used for evaluation. The Kalman filter was used for a comparison with the CSNNRF. Figures 3(a)-(c) show the analysis results.

As shown in Fig. 3(a), for comfortable sound, a low-pass filter was evaluated most highly in all ranges of SN ratios. It was found by the t-test ($p<0.05$) that the CSNNRF received a better evaluation than the Kalman filter. In addition, the CSNNRF received a significantly higher evaluation than non-processing when the SN ratio was between 0.8[dB] and 5.8[dB]. Although Evaluation value seems less than non-processing in the higher range of the SN ratio, there was no significant difference based on the results of the t-test.

As shown in Fig. 3(b), regarding clarity, it was found by the t-test ($p<0.05$) that the CSNNRF

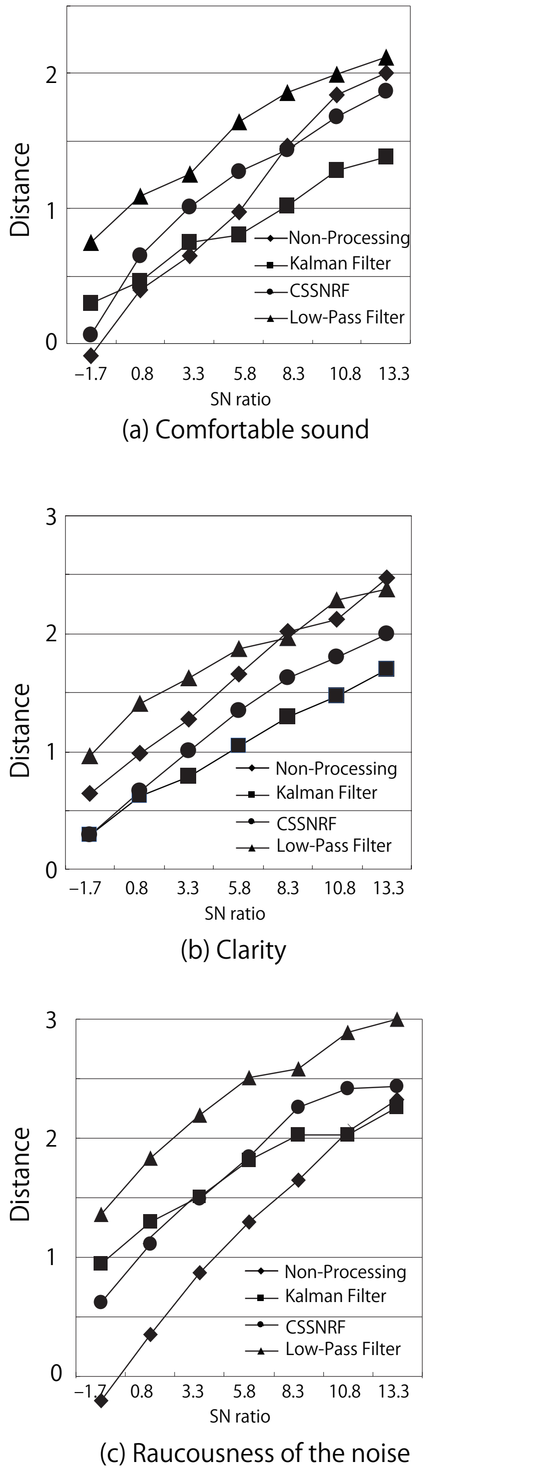

Figure3 Results of subjective evaluation of the CSNNRF

received a better evaluation than the Kalman filter in all ranges of the SN ratio. However non-processing or the low-pass filter showed greater clarity in comparison with the CSNNRF. It is likely that processing by the CSNNRF or Kalman filter distorted the speech signal and then lowered its clarity.

The raucousness of the noise was significantly improved with the CSNNRF ($p<0.05$). As shown in Fig. 3(c), it was found by the t-test ($p<0.05$) that the CSNNRF received a better evaluation than the Kalman filter, but a poorer evaluation than the low-pass filter. It must be ascribed to a high-frequency noise with a varying intensity that was left after CSNNF.

5. Conclusion

In this paper, we developed a new adaptive noise reduction filter (CSNNFR) which expanded the SNNRF. The CSNNRF improved the SN ratio by about 1.0-6.2 [dB]. The lower the SN ratio of the input signal was, the higher the improvement rate was. The SN ratio improvement rate with the CSNNRF was superior to that of the Kalman filter.

On evaluation in the hearing experiment, the low-pass filter was effective although its improvement rate of the SN ratio was markedly lower than the CSNNRF. The reason is likely that humans generally ignore noise of a constant low frequency. A problem still remains with the CSNNRF in terms of "Clarity". On the other hand, it was shown that the CSNNRF perform better than the Kalman filter regarding "Comfortable sound" and "Raucousness of the noise".

References

[1] Itsuo Ono, " Today's digitally hearing aid technology and outlook for the future", The Acoustical Society of Japan, Vol.47, No.10, 778–784, 1991.

[2]Koichi Yanagisawa, Kayoko Tanaka, Itsuo Yamaura," Noise Reduction of a Digital Hearing Aid for Continuous Speech Based on Spectral Harmonic Structure", Electrical Engineering in Japan, Vol.120-C, No.3, 382–389, 2000.

[3]Yuichi Ueda, Takayuki Agawa, Masazumi Aoyama, Akira Watanabe," Amplitude Compression Method for a Digital Hearing Aid Using a Composite Filter", IEICE TRANS., Vol. J81–A, No12, 1728–1738, 1998.

[4] Hiroki YOSHIMURA, Tadaaki SHIMIZU, Naoki ISU, Kazuhiro SUGATA,"Construction of Noise Reduction Filter by Use of Sandglass-Type Neurl Network " , IEICE TRANS. FUNDAMENTALS , Vol. E80-A, No.8, 1384–1390, 1997.

[5]Hiroki YOSHIMURA, Tadaaki SHIMIZU, Naoki ISU, Kazuhiro SUGATA," An Adaptive Noise Reduction Filter for Discrete Signal by Use of Sandglass-Type Neural Network", Electrical Enginering in Japan, Vol 127, No.4, 39–50,1999.

[6]Hiroki YOSHIMURA, Tadaaki SHIMIZU, Naoki ISU, Kazuhiro SUGATA,"Adaptive Noise reduction by Using Cascaded Sandglass-Type Neural Network", Electrical Engineering in Japan, Vol 121-C, No.2, 681–688,2001.