Noise Suppression Algorithm Based on Loudness Management for Preserving Speech Components

Nobuhiko Hiruma[1], *Hidetoshi Nakashima*[2], *Yoh-ichi Fujisaka*[1]

[1]Rion Co., Ltd., Japan
[2]National Institute of Technology, Kumamoto College, Japan

n-hiruma@rion.co.jp, nakashi@kumamoto-nct.ac.jp, y-fujisaka@rion.co.jp

1. Introduction

A hearing aid amplifies the input sound from a microphone in conformity with the hearing threshold of a user; however, it simultaneously enhances undesired environmental sounds such as the noise of an air conditioner or a car. According to MarkeTrak V [1], such kinds of noise are often perceived by the hearing aid user with a strong feeling of discomfort, and thereby, the user perceives the continuous use of the hearing aid to be inconvenient.

Numerous algorithms have been proposed and implemented for hearing aids as well as cell phones for the purposes of speech enhancement, reduction of discomfort due to noise [2, 3], and alleviation of listening effort [4]. However, it is widely known today that speech intelligibility is not significantly improved by using conventional algorithms, despite the improvement in sound quality [5]. Considering that a hearing aid is an effective communication device for the hearing-impaired, the most important issue for users in their daily lives is to improve the speech recognition rate during communication. Noise reduction algorithms based on spectral subtraction estimate the spectrum of the noise component and then subtract it from the observed signal spectrum; even hence, if the speech spectrum is masked in the noise, it would result in a decrease in the output signal or level of loudness. The degradation of output level for certain frequency components might affect the speech quality depending on the hearing level of the patients, because the output sound level will be lower than the patients threshold of hearing, though this degradation has a positive effect in terms of reducing the annoyance of patients.

In this study, a noise reduction algorithm based on loudness management, which eliminates unnecessary spectral subtraction, is proposed. The concept of our algorithm is that noise components are preserved as much as possible by keeping loudness level of speech signal for positive effect of speech recognition.

2. Proposed System

Figure 1 represents a block diagram of our proposed system. Here, the observed signal $x(t)$ is defined with speech signal $s(t)$ and noise $n(t)$ as

$$x(t) = s(t) + n(t) \tag{1}$$

where t denotes the time index. The above Eq.1 can be written in the frequency domain as

$$X(\omega) = S(\omega) + N(\omega), \tag{2}$$

where ω denotes the angular frequency. The estimated speech signal spectrum is calculated using estimated noise power [6] $|N'(t,\omega)|$, as

$$|S'(t,\omega)|^2 = |X(t,\omega)|^2 - |N'(t,\omega)|^2. \tag{3}$$

For the calculation of partial loudness based on ISO 532B, the observed and estimated speech signal spectra, $|X(t,\omega)|^2$ and $|S'(t,\omega)|^2$, are transformed into one-third octave band energy, $E_x(t,m)$ and $E_{s'}(t,m)$, respectively. Here, m is one-third octave band index.

The partial loudness of the observed signal, $L_x(t,k)$, estimated speech signal, $L_{s'}(t,k)$, and output signal, $L_y(t,k)$, are calculated based on DIN45631 [7] using $E_x(t,m)$ and $E_{s'}(t,m)$. Here, k denotes bark band index and the unit of the partial loudness is sone. In this algorithm, noise suppression gain is determined to make $L_y(t,k)$ correspond to $L_{s'}(t,k)$. Hence, the difference between $L_x(t,k)$ and $L_{s'}(t,k)$ is obtained as

$$L_d(t,k) = L_x(t,k) - L_{s'}(t,k), \tag{4}$$

where $L_d(t,k)$ means suppression gain in sone for the correspondence between $L_y(t,k)$ and $L_{s'}(t,k)$. The suppression gain $G(t,w)$ is obtained by using $L_d(t,k)$ and look-up table for conversion from sone to dB, and then it is transformed into time domain to get FIR filter coefficient $w(t)$ by inverse fast Fourier transform (IFFT). The output signal $y(t)$ is obtained by convolution of $x(t)$ and $w(t)$.

3. Simulation

To confirm the feasibility of the proposed system, a numerical simulation was conducted using MATLAB. The temporal variation of the partial loudness level of the third bark is shown in Figure 2. In this figure, the blue, black, and red lines represent partial loudness of the observed signal $L_x(t,k)$, true speech $L_s(t,k)$, and estimated speech $L_{s'}(t,k)$. Figure 3 shows the result of instantaneous partial loudness level in each bark. In this figure, the green line denotes the partial loudness level of output signal $L_y(t,k)$. It appears that the black and red lines are almost similar. These results indicate that the partial loudness level of estimated speech is almost the same as that of true speech.

4. Evaluation

Evaluation of the output signal in the proposed system is performed using the short-time objective intelligibility (STOI) [8]. For the sake of comparison, the output of conventional noise reduction is also evaluated. In this evaluation, pink noise is added and its level varies from -10 dB to 10 dB of signal-to-noise ratio (SNR) in steps of 2 dB. Figure 4 shows the result of this simulation. The red and blue lines represent the score of output signal of the proposed and conventional methods, respectively.

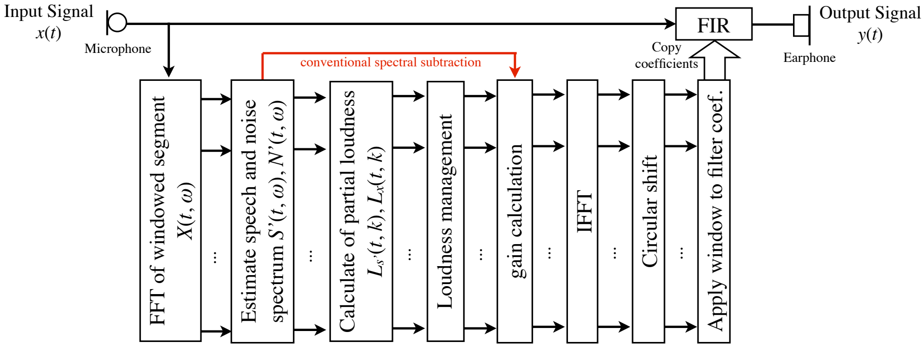

Figure 1: *Block diagram of proposed system.*

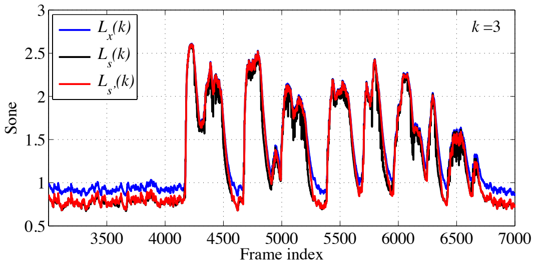

Figure 2: *Temporal variation of the partial loudness levels.*

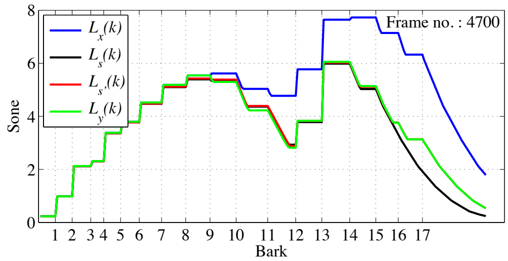

Figure 3: *Instantaneous partial loudness levels.*

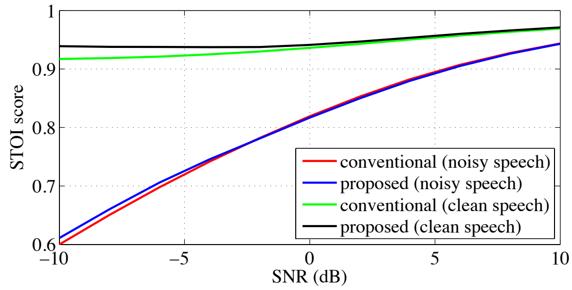

Figure 4: *SNR vs. STOI.*

The green and black lines represent the score of speech component of the output signal by applying the gain function to these two methods, which is obtained by the convolution of speech signal with the gain function in the proposed and conventional methods.

For all the investigated SNR conditions, the trend of STOI scores for the proposed method is almost the same as that for the conventional method, although the proposed method included extra noise, which is expected to include weak speech spectral components. On the other hand, the scores of the speech part with the gain function of the proposed method are superior to that of the conventional method for low SNR conditions. This result reveals that the proposed method has gain control based on partial loudness, and improves the speech signal quality effectively, while including the extra noise. We think that there is some possibility about positive effect of speech recognition by using the extra noise, which has the same intonation as the estimated speech [9].

5. Conclusions

In this paper, a noise suppression algorithm based on loudness management is proposed. The results of our evaluation using STOI indicate that the score of the clean-speech output by applying the gain function in the proposed method is improved compared with that of the conventional method, while the score of noisy-speech by the proposed method is almost the same as that of the conventional method. These results imply that speech components are preserved; however, a subjective evaluation has not been performed.

6. References

[1] S. Kochkin, "MarkeTrack V: Why my hearing aids are in the drawer: The consumers' perspective," *The hearing journal*, vol.53, pp. 34-41, 2000.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. 27, no. 2, pp. 113-120, 1979.

[3] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. on Acoust., Speech, and Signal Processing*. vol. 32, no. 2, pp. 1109-1121, 1984.

[4] A. Sarampalis, S. Kalluri,B. Edwars and E. Hafler, "Objective measures of listening effort: effects of background noise and noise reduction," *J. Speech Lang Hear Res*. vol. 52, no. 5, pp. 1230-1240, 2009.

[5] P. C. Loizou, and G. Kim, "Reasons why Current Speech- Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions," *IEEE Trans. on Audio, Speech and Language Processing*. vol. 19, no. 1, pp. 47-56, 2011.

[6] S. Rangachari, and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *IEEE Trans. on Acoust., Speech, and Signal Processing*. vol. 48, no. 2, pp. 220-231, 2006.

[7] E. Zwicker, H. Fastl ,U. Widmann, K. Kurakata, S. Kuwano, and S. Noamba, "Program for calculating loudness according to DIN 45631 (ISO532B)," *Journal of the Acoustical Society of Japan*. vol. 12, no. 1, pp. 39-42, 1991.

[8] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," *IEEE Trans. on Audio, Speech, and Language Processing*. vol. 19, no. 7, pp. 2125-2136, 2011.

[9] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wyhonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*. vol. 270, no. 5234, pp. 303-304, 1995.