Blind Reverberation Power Estimation Using Exponential Averaging with Attack and Release Time Constants for Hearing Aids

Kotoyo Nozaki[1*], *Yasuhiro Oikawa*[1**], *Yusuke Ikeda*[2], *Yoh-ichi Fujisaka*[3], *Masahiro Sunohara*[3]

[1]Waseda University, Japan
[2]Tokyo Denki University, Japan
[3]RION Co., Ltd., Japan

*kotokoto3726@asagi.waseda.jp, **yoikawa@waseda.jp

Abstract

Dereverberation processing is necessary for hearing-aid systems with a limitation of computational cost because reverberation degrades speech intelligibility in some reverberation environments. The spectral subtraction (SS) method is a simple and well-known technique as not only a noise-reduction method but also a dereverberation method. In dereverberation methods that are based on SS for hearing aids, it is desirable to estimate the reverberation power with blind processing. The SS-based blind-estimation method was proposed by Sunohara et al. using exponential averaging with attack and release time constants for single-channel speech signals.

In this paper, we evaluate the estimation accuracy of the reverberation components of the proposed method. The estimation error, which is the difference between the true and estimated reverberation power was used for evaluation, that was compared the results obtained with the method proposed by Lebart et al., which is a non-blind SS-based dereverberation method. From the results, the reverberation power was more correctly estimated, especially in the case of long reverberation, and the estimation error of the proposed method was about a half of the well-known non-blind method by Lebart.

Index Terms: reverberant speech, blind estimation, hearing aids, exponential averaging, spectral subtraction

1. Introduction

Speech recognition is difficult owing to the reverberation. In particular, hearing-impaired persons have difficulties when listening to a speech in reverberation environments. For example, their recognition rates of speech decrease as the reverberation time increases [1]. For severely hearing-impaired listeners, it becomes hard to follow conversations in noisy or reverberant conditions even when speaking with only one person, as Moore et al. reported [2].

According to Folkeard et al.[3], listening may be improved performing dereverberation processing in hearing aids for hearing-impaired persons. Many researchers have studied dereverberation processing for a long time, and many methods have been proposed. Neely et al. proposed a method using inverse filtering of the room impulse response [4], and Gannot et al. employed a signal subspace approach [5]. Although these methods are effective for dereverberation, it is hard for them to be implemented into hearing aids because of the limited of computational resource available. A method that is considered to be relatively simple is the spectral subtraction (SS) method [6], which is effective not only for noise reduction but also for dereverberation [7]. Löllmann proposed a dereverberation method based on the SS method for hearing-aid systems[8].

The accurate estimation of reverberant components is important for the SS method. There are two kinds of estimation methods—namely, non-blind methods and blind methods. For non-blind estimation methods [9, 10], information about the sound field, such as the reverberation time, is required beforehand. On the other hand, ordinary blind estimation methods do not require previous information about the sound field [11, 12, 13]. However, most of them require multiple microphones to estimate the reverberant components; hence, it is difficult to implement them into hearing aids. Recently, we proposed a dereverberation system with the blind estimation method using only single-channel speech signals based on exponential averaging with attack and release time constants [14, 15].

In this paper, we evaluate our proposed dereverberation system with blind estimation of reverberation power was evaluated. True reverberation power was calculated from the impulse responses without direct sound to obtain an estimation error, which is the difference between true and estimated reverberation powers. The estimation error of the proposed estimation method was compared with that of a typical non-blind method proposed by Lebart [9].

2. Dereverberation Method

2.1. Spectral Subtraction

The SS method was proposed by Boll in 1979 as a noise-reduction method [6]. It is also known to be valid for dereverberation. Then, Kinoshita et al. evaluated the validity of the SS method for the reduction of latter reverberant components caused by evaluation experiments using spectrogram and automatic speech recognition (ASR) [7]. The above-mentioned method is relatively simple compared with other methods, such as inverse filtering or the signal subspace approach [4, 5]. The procedure of the SS method is as follows:

Step:1 Perform a Fourier transformation on the input signal.

Step:2 Estimate the noise/reverberation power of the input signal.

Step:3 Subtract the estimated noise/reverberation power from the power of the input signal.
(This subtracted power is treated as the gain.)

Step:4 Multiply the gain by the input power.

In the second step in the procedure, it is necessary to accurately estimate the reverberation power for dereverberation. In the dereverberation method proposed by Lebart, the reverberation power is estimated using an impulse response model [9], and it is a well-known non-blind estimation method.

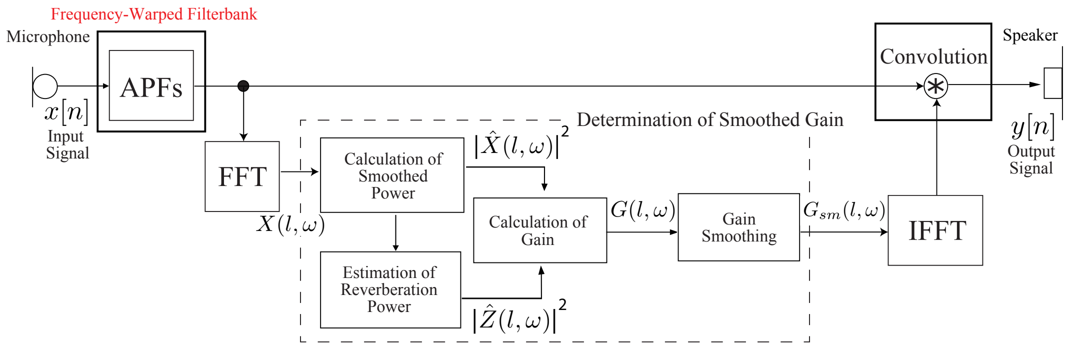

Figure 1: *System overview. The input signal flows through all-pass filters (APFs) with the warping parameter to analyze the input signal. The gain function $G(l,\omega)$ is determined by the estimated reverberation power.*

2.2. Exponential Averaging with Attack and Release Time Constants

Figure 1 shows the signal flow of our dereverberation system, which includes our proposed estimation method for reverberation power. An input signal from the microphone is analyzed using a frequency-warped filterbank (FWF) [16, 17], and transformed into the frequency domain using FFT. Sunohara et al. applied the weighted overlap-add (WOLA) filterbank[18] in a part of analysis/synthesis[14]. After the analysis, the smoothed signal power is calculated, and then the reverberation power is also estimated from the smoothed signal power. The gain function, which is derived from the above two powers, was smoothed in order to avoid musical noise issues in the output signal. The smoothed gain function was then transformed into the time domain and convolved with the warped signal that was passed through in each all-pass filter (APF).

Exponential averaging is one of the smoothing methods that were proposed by Roberts [19], and it is frequently used as a low-pass filter (LPF) on the field of signal processing. The weighting coefficient changes depending on the relationship between the estimated reverberation power and the input signal power. This method is given as follows.

Input signal $x(t)$ is generated by convolving the source signal $s(t)$ with the room impulse response $h(t)$, which includes the reverberation component,

$$x(t) = s(t) * h(t), \tag{1}$$

where $*$ represents the convolution operation. The frequency-analyzed input signal obtained by FWF is represented as $X(l,\omega)$, where l and ω denote the time and frequency indexes, respectively. The smoothed power $|\hat{X}(l,\omega)|^2$ is given as

$$|\hat{X}(l,\omega)|^2 = \beta|X(l,\omega)|^2 + (1-\beta)|\hat{X}(l-1,\omega)|^2, \tag{2}$$

where the parameter $\beta\,(0 < \beta \leq 1)$ represents the weighting value in exponential averaging. $|\hat{X}(l,\omega)|^2$ includes the reverberation power $|\hat{Z}(l,\omega)|^2$. Sunohara et al. estimated the reverberation power [14],

$$|\hat{Z}(l,\omega)|^2 = \gamma(l)|\hat{X}(l,\omega)|^2 + \{1-\gamma(l)\}|\hat{Z}(l-1,\omega)|^2, \tag{3}$$

where $\gamma(l)$ is a parameter that determines the time constant of exponential averaging

$$\gamma(l) = \begin{cases} \gamma_{\text{at}}, & (|\hat{X}(l,\omega)|^2 > |\hat{Z}(l-1,\omega)|^2) \\ \gamma_{\text{re}}, & (\text{otherwise}) \end{cases}. \tag{4}$$

Finally, the gain function $G(l,\omega)$ is obtained from the smoothed power of the input signal $|\hat{X}(l,\omega)|^2$ and reverberation signal $|\hat{Z}(l,\omega)|^2$,

$$G(l,\omega) = \begin{cases} \sqrt{\frac{|\hat{X}(l,\omega)|^2 - |\hat{Z}(l,\omega)|^2}{|\hat{X}(l,\omega)|^2}} & (|\hat{X}(l,\omega)|^2 > |\hat{Z}(l-1,\omega)|^2) \\ 0 & (\text{otherwise}) \end{cases}. \tag{5}$$

Then, the gain function $G(l,\omega)$ is smoothed to $G_{sm}(l,\omega)$ using the exponential averaging, as shown in Eq.(2). $G_{sm}(l,\omega)$ is transformed by IFFT and convolved with the signal analyzed by APFs as shown in Fig.1.

2.3. Well-known non-blind estimation method

Lebart used the power spectral density (PSD) to estimate the reverberation power from the reverberation time, which is a well-known non-blind estimation method [9].

$$|\tilde{Z}(l,\omega)|^2 = e^{-2\Delta l_0}|\hat{X}(l-l_0,\omega)|^2, \tag{6}$$

where l_0 is the typical duration for which it can be assumed that the signal is stationary. In this paper, let l_0 be 50 ms and

$$\Delta = \frac{3\ln 10}{T_r}, \tag{7}$$

where T_r is the reverberation time.

3. Experiment

In this section, a numerical experiment is conducted in order to evaluate the accuracy of proposed method by comparing the estimated reverberation power with true one as the estimation error.

3.1. Method

The reverberation speeches for the experiment were created by convolving anechoic speeches with an impulse response. The reverberation was modeled based on Polack's model [20]. In Polack's model, it is assumed that the room impulse response $h(t)$ is generated using an unsteady stochastic process

$$h(t) = \begin{cases} 0, & t < 0 \\ b(t)e^{-\Delta t}, & t \geq 0 \end{cases}, \tag{8}$$

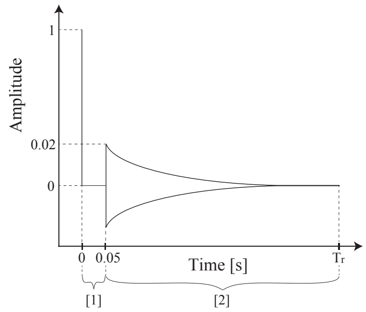

Figure 2: *Pattern diagram of the room impulse response based on the model proposed by Polack [20]. The first part $h_d(t)$ represents a direct sound and the second part $h_r(t)$ represents a reverberation component. The border between the direct sound and the reverberation component was 50 ms.*

where $b(t)$ is a steady Gaussian noise with an average of 0, and Δ is shown in Eq.(7).

Impulse responses were created based on this model. A direct sound is assigned at the beginning of the impulse response, and the reverberation decay starts from 50 ms after the direct sound. Let $h_d(t)$ be the first part before 50 ms and $h_r(t)$ be the second part after 50 ms. The impulse response with direct sound and the reverberation component is defined by

$$\hat{h}(t) = \begin{cases} 0, & t < 0 \\ h_d(t), & 0 \leq t < 50 \text{ ms} \\ h_r(t), & 50 \text{ ms} \geq t \end{cases}, \tag{9}$$

$$h_d(t) = \begin{cases} 1, & t = 0 \\ 0, & 0 \geq t \end{cases}, \tag{10}$$

$$h_r(t) = b(t)e^{-\Delta t}, \tag{11}$$

where $b(t)$ is the same as shown in Eq. (8). Figure 2 shows a pattern diagram of the impulse response.

For the evaluation, we used the difference between the estimated reverberation power and the true reverberation power, which is the estimation error. First, a true reverberation speech without direct sound was generated by convolving an anechoic speech with $h_r(t)$,

$$z_{true}(t) = s(t) * h_r(t), \tag{12}$$

Table 1: *Experimental condition.*

Parameter	**Value**
γ_{at} (time constant)	3.125×10^{-5} (2.0 s)
γ_{re} (time constant)	6.250×10^{-4} (1.0×10^{-1} s)
Gain estimation parameter: attack (time constant)	1.250×10^{-2} (5.0×10^{-3} s)
Gain estimation parameter: release (time constant)	9.375×10^{-4} (6.7×10^{-2} s)
Smoothing parameter: β	1.250×10^{-3}
Lower bound of gain reduction: η	0
Sampling frequency [kHz]	16
Reverberation time [s]	$1.0, 1.5, 2.0$
FFT length [samples] (time)	32 (2 ms)

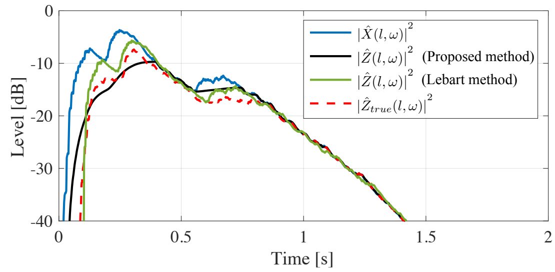

Figure 3: *Level of estimated powers. This figure shows the level of powers of the input signal (blue line), the reverberation signals estimated by the proposed method (black line) and by the non-blind method proposed by Lebart (green line), and the true reverberation signal (red broken line) at the 8th frequency band in FWF, whose central frequency was 1100 Hz. The reverberation time T_r was 1.5 s.*

which is $z_{true}(t)$ a true reverberation speech. It was transformed to $|Z_{true}(l,\omega)|$ in the frequency domain by FFT, and $|Z_{true}(l,\omega)|$ was smoothed according to Eq.(2),

$$|\hat{Z}_{true}(l,\omega)|^2 = \beta|Z_{true}(l,\omega)|^2 + (1-\beta)|\hat{Z}_{true}(l-1,\omega)|^2. \tag{13}$$

Second, the sum of the estimated reverberation power $|\hat{Z}(l,\omega)|^2$ and the true reverberation power $|\hat{Z}_{true}(l,\omega)|^2$ was calculated for each method, after which the sum of the difference between the estimated and true reverberation powers $S_{diff}(\omega)$ was calculated,

$$S_{est}(\omega) = \sum_l |\hat{Z}(l,\omega)|^2, \tag{14}$$

$$S_{true}(\omega) = \sum_l |\hat{Z}_{true}(l,\omega)|^2, \tag{15}$$

$$S_{diff}(\omega) = \sum_l ||\hat{Z}(l,\omega)|^2 - |\hat{Z}_{true}(l,\omega)|^2|. \tag{16}$$

In this analysis, the summation was calculated for 3.0 s from the beginning. Finally, the ratio of the sum of the difference power $S_{diff}(\omega)$ and the sum of true reverberation power $S_{true}(\omega)$ was defined as the error ratio in the reverberation power estimation $Z_{ER}(\omega)$,

$$Z_{ER}(\omega) = \frac{S_{diff}}{S_{true}}. \tag{17}$$

We compared the values of Z_{ER} of the proposed method and the well-known Lebart method at each reverberation time. The speech sources were selected from the familiarity-controlled word-lists (FW03) [21], which contains words spoken by 2 male and 2 female Japanese speakers. In this experiment, 100 words were used for each speaker, and the total number of words was 400. After calculating $Z_{ER}(\omega)$ for each sound, histograms were approximated with the kernel distribution using MATLAB, and the expected value $E(\omega)$ was then calculated for each frequency band. The number of bins on the histograms was 100, and the parameters are defined in Table 1.

3.2. Results

Figure 3 represents one of the estimated powers when the reverberation time T_r was 1.5 s. This figure shows each power of the input signal, reverberation signals estimated by the proposed method and Lebart method, and the true reverberation signal. The power of the reverberation signal estimated by the Lebart method is larger than that of the reverberation signal estimated by our proposed method, especially at the beginning part. It is also larger than the power of the true reverberation signal.

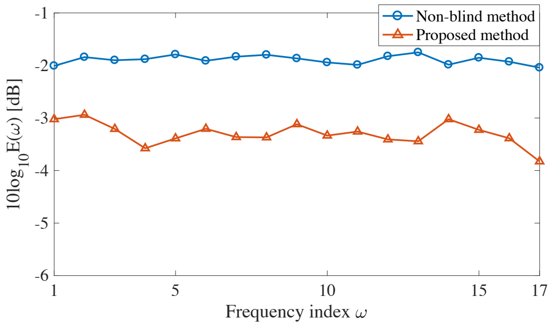

Figure 4: *Experimental results obtained when the reverberation time T_r was 1.0 s. The vertical axis represents $E(\omega)$ of the logarithmic scale and the horizontal axis represents the frequency index ω in FWF.*

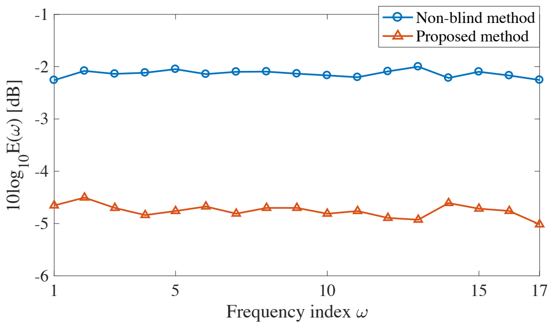

Figure 5: *Experimental results obtained when the reverberation time T_r was 1.5 s. The vertical axis represents $E(\omega)$ of the logarithmic scale and the horizontal axis represents the frequency index ω in FWF.*

Figure 4, 5 and 6 show the error ratio $E(\omega)$ of the logarithmic scale at each reverberation time, 1.0 s, 1.5 s, and 2.0 s, respectively. In each figure, the blue line shows $E(\omega)$ in the non-blind method, and the orange line shows $E(\omega)$ in the proposed method. For the proposed method, $E(\omega)$ exceeded that of the non-blind method proposed by Lebart at each reverberation time and in each frequency band. In the case of long reverberation condition, the value of $E(\omega)$ for the proposed method was around 3 dB lower than that of the non-blind method proposed by Lebart. This means that in the proposed method, the value of $E(\omega)$ was about one half of Lebart's method. It was also shown that by using our proposed method, the reverberation power is more correctly estimated than the other method. Because the value of $E(\omega)$ in the proposed method decreases as the reverberation time increases, the proposed method may be more effective as the reverberation time increases.

4. Conclusions

We evaluate the implementation of our newly proposed blind reverberation power-estimation method which was proposed for the purpose of implementing that into hearing-aid devices. The estimated reverberation power of the proposed method was

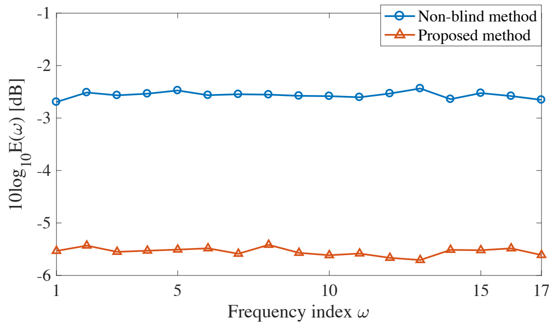

Figure 6: *Experimental results obtained when the reverberation time T_r was 2.0 s. The vertical axis represents $E(\omega)$ of the logarithmic scale and the horizontal axis represents the frequency index ω in FWF.*

compared with that of well-known non-blind methods. We performed experiments for three kinds of reverberation times. The results obtained suggested that the proposed method is more effective than the well-known non-blind method, regardless of the reverberation time. In the future, the speech signal processed by this proposed method will be evaluated using objective/subjective experiments. Furthermore, reverberation power-estimation methods for impulse responses including early reflection sounds will be studied.

5. References

[1] A. K. Nabelek and J. M. Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," *J. Speech, Lang., and Hear. Res.*, vol. 17, no. 4, pp. 724–739, 1974. [Online]. Available: + http://dx.doi.org/10.1044/jshr.1704.724

[2] B. C. J. Moore, *Cochlear hearing loss*. John Wiley and Sons, 2007.

[3] L. V. S. S. Folkeard, P, "Using a de-reverberation program to improve speech intelligibility and reduce perceived listening effort," *Hearing Review*, vol. 24, no. 4, pp. 32–33, 2017.

[4] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, 1979.

[5] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP J. Appl. Signal Process.*, vol. 2003, pp. 1074–1090, Jan 2003. [Online]. Available: http://dx.doi.org/10.1155/S1110865703305049

[6] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr 1979.

[7] K. Kinoshita, T. Nakatani, and M. Miyoshi, "Efficient blind dereverberation framework for automatic speech recognition," in *INTERSPEECH*, 2005.

[8] H. W. Löllmann and P. Vary, "Low delay noise reduction and dereverberation for hearing aids," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, p. 437807, 2009.

[9] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica united Acustica*, vol. 87, no. 3, pp. 359–366, 2001.

[10] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," 2007.

[11] K. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 5, pp. 1579–1591, 2007.

[12] L. Wang, K. Odani, and A. Kai, "Dereverberation and denoising based on generalized spectral subtraction by multi-channel LMS algorithm using a small-scale microphone array," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 12, 2012.

[13] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 7, pp. 1732–1745, 2010.

[14] M. Sunohara, M. Nakaichi, and Y. Kondou, "Simple dereverberation method for hearing aid users (in Japanese)," *Meet. Acoust. Soc. Jpn.*, pp. 853–854, 2015.

[15] K. Nozaki, Y. Ikeda, Y. Oikawa, Y. Fujisaka, and M. Sunohara, "Low latency dereverberation with considering speech naturalness for hearing aids," *Joint Meet. Acoust. Soc. Am. Acoust. Soc. Jpn.*, 2016.

[16] A. Harma, M. Karjalainen, L. Savioja, V. Valimaki, U. K. Laine, and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *J. Audio. Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, 2000.

[17] J. M. Kates and K. H. Arehart, "Multichannel dynamic-range compression using digital frequency warping," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 3003–3014, Jan 2005. [Online]. Available: http://dx.doi.org/10.1155/ASP.2005.3003

[18] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*, ser. Prentice-Hall Signal Processing Series: Advanced monographs. Prentice-Hall, 1983. [Online]. Available: https://books.google.co.jp/books?id=X_NSAAAAMAAJ

[19] S. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.

[20] J. D. Polack, "La transmission de l'énergie sonore dans les salles," Ph.D. dissertation, 1988. [Online]. Available: http://www.theses.fr/1988LEMA1011

[21] S. Amano, K. Kondo, Y. Suzuki, and S. Sakamoto, "Speech data set for word intelligibility test based on word familiarity (FW03)," *NII Speech Resources Consortium*, 2006.