14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

ON ROBUST INVERSE FILTER DESIGN FOR ROOM TRANSFER FUNCTION FLUCTUATIONS

Takafumi Hikichi[†], *Marc Delcroix*[†‡] *and Masato Miyoshi*[†‡]

[†]NTT Communication Science Laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, "Keihanna Science City", Kyoto, 619-0237, Japan
[‡]Graduate School of Information Science and Technology, Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo, 060-0814 Japan
email: {hikichi,marc.delcroix,miyo}@cslab.kecl.ntt.co.jp

ABSTRACT

Dereverberation methods based on the inverse filtering of room transfer functions (RTFs) are attractive because high deconvolution performance can be achieved. Although many methods assume that the RTFs are time-invariant, this assumption would not be guaranteed in practice. This paper deals with the problem of the sensitivity of a dereverberation algorithm based on inverse filtering. We evaluate the effect of RTF fluctuations caused by source position changes on the dereverberation performance. We focus on the filter energy with a view to making the filter less sensitive as regards these fluctuations. By adjusting three design parameters, namely, filter length, modeling delay, and regularization parameter, a dereverberation performance of up to 15 dB of the signal-to-distortion ratio could be obtained when the source position was changed with one-eighth wavelength distance, whereas conventional investigations have claimed that such a variation would cause a large degradation.

1. INTRODUCTION

The deconvolution of room transfer functions (RTFs) is useful in various applications such as sound reproduction, sound field equalization, and speech dereverberation. Usually, RTFs are modeled as linear time-invariant systems, and estimated as FIR filters. Deconvolution is performed by using inverse filters that are designed to equalize these FIR filters. However, this time-invariance is not always guaranteed in realistic situations. For example, RTFs vary with changes in the source and/or receiver positions, temperature and other environmental factors [1, 2, 3]. As a result, an inverse filter correctly designed for one condition may not work well for another condition, and some kind of compensation or adaptation is necessary.

Robustness issue of sound equalization in relation to the movement of a sound source or receiver has been addressed in several papers. In [4, 5], the mean squared error caused by movement of the source or receiver is derived based on statistical room acoustics. These studies show that even a small movement of a few tenths of a wavelength degrades the equalization performance. Although previous studies have claimed that the technique is sensitive to small RTF changes, we believe it can be improved. The focus of these studies was to investigate the influence of the source or microphone changes on the performance of a fixed inverse filter, and not on how to design a more robust inverse filter.

The purpose of this paper is to pursue ways of designing inverse filters that would be less sensitive to RTF variations. We consider that reducing the filter energy is the key to making the filter less sensitive, since less degradation is expected if a filter with a smaller energy can be used. From this point of view, we focus on the influence of three parameters used in the design of inverse filters: filter length, modeling delay, and a regularization parameter. By selecting proper parameter values, we expect to make the filter more robust to RTF variations. We attempted to find adquate values for these three parameters by investigating their influences on the performance.

In this paper, we deal with a single-source and multiple-microphone acoustic system. The following section describes the dereverberation algorithm, and then analyzes the effect of the three design parameters on the filter norm.

2. DEREVERBERATION ALGORITHM AND DESIGN PARAMETERS

2.1 Dereverberation algorithm

The dereverberation algorithm used in this study is proposed in [6], and is based on the channel estimation technique and the Multiple input/output INverse Theorem (MINT) [7]. The channel estimation principle is based on the cross-relation of the multi-channel microphone signals. With two channels, impulse response estimates can be obtained through the eigenvalue decomposition of the data correlation matrix [8, 9]. Then, the inverse filter set is calculated using those multiple impulse response estimates. The inverse filter vector, denoted as $\mathbf{g}$, satisfies the following equation,

$$\mathbf{H}\mathbf{g} = \mathbf{v}, \tag{1}$$

where

$$\mathbf{H} = [\mathbf{H}_1, \cdots, \mathbf{H}_P],$$

$$\mathbf{H}_i = \underbrace{\left.\begin{pmatrix} h_i(0) & 0 & \dots & 0 \\ h_i(1) & h_i(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ h_i(J) & & & h_i(0) \\ 0 & h_i(J) & & h_i(1) \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & h_i(J) \end{pmatrix}\right\}}_{M} (J+M), i = 1, \cdots, P,$$

$$\mathbf{g} = [\underbrace{g_1(1),\ldots,g_1(M),\cdots,\ g_P(1),\ldots,g_P(M)}_{PM}]^T,$$
$$\mathbf{v} = [\underbrace{0,\ldots,0}_{d},1,0,\ldots,0]^T,$$

P is the number of channels, $h_i(n)$ is the impulse response estimate between the source and the i-th microphone, J is the number of taps of the impulse response estimates, $\mathbf{g}$ is the inverse filter vector, M is the filter length for each channel, and d is the modeling delay. An arbitrary delay can be inserted in the equalized response by setting d ($d \geq 0$). Hereafter, we assume that matrix $\mathbf{H}$ is full row rank, and consider that the impulse response estimates are normalized.

The inverse filter vector can be obtained by

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{v}, \tag{2}$$

where $\mathbf{A}^{+}$ is the Moore-Penrose pseudo inverse of matrix $\mathbf{A}$. An inverse filter with the minimum length is calculated by setting M so that matrix $\mathbf{H}$ is square, i.e., $(J+M) = PM$ holds, which leads to $M = J/(P-1)$. Filter length can be set at $M > J/(P-1)$ as well.

2.2 Design parameters

In this section, we study the influence of three design parameters in the inverse filter calculation. The following analysis shows that we can expect the regularization parameter, filter length, and modeling delay to be effective in reducing the filter norm and hence increasing the robustness against RTF variations.

2.2.1 *Regularization*

Cost function C for designing an inverse filter with regularization is expressed as,

$$C = ||\mathbf{v} - \mathbf{H}\mathbf{g}||^2 + \delta||\mathbf{g}||^2. \tag{3}$$

where $\delta(\geq 0)$ is called the regularization parameter. The solution that minimizes the cost function expressed in Eq. (3) is given by the following [10].

$$\mathbf{g}(\delta) = (\mathbf{H}^T\mathbf{H} + \delta\mathbf{I})^{-1}\mathbf{H}^T\mathbf{v}, \tag{4}$$

where $\mathbf{I}$ is an identity matrix. The power of the L2-norm of this filter vector becomes,

$$\begin{aligned}
||\mathbf{g}(\delta)||^2 &\leq ||(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}\mathbf{H}^T)||^2 \\
&= ||\mathbf{H}(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}\mathbf{H}^T|| \\
&= ||\mathbf{H}^T\mathbf{H}(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}|| \\
&\approx ||(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}|| \qquad (5) \\
&= 1/\mu_{min}[(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})] \qquad (6) \\
&\leq ||\mathbf{g}||^2.
\end{aligned}$$

Here, in the derivation of the approximation in Eq.(5), we apply the Taylor expansion to the term $(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1}$. Let $(\mathbf{H}^T\mathbf{H})^{-1}$ denote $\mathbf{G}$. Assuming that δ is sufficiently small, $||\mathbf{I}|| \gg ||\delta\mathbf{G}||$ holds. Then,

$$\begin{aligned}
(\mathbf{H}^T\mathbf{H}+\delta\mathbf{I})^{-1} &= ((\mathbf{H}^T\mathbf{H})(\mathbf{I}+\delta(\mathbf{H}^T\mathbf{H})^{-1}))^{-1} \\
&= ((\mathbf{H}^T\mathbf{H})(\mathbf{I}+\delta\mathbf{G}))^{-1} \\
&= (\mathbf{I}+\delta\mathbf{G})^{-1}(\mathbf{H}^T\mathbf{H})^{-1} \\
&= (\mathbf{I}-\delta\mathbf{G}+\delta^2\mathbf{G}^2-\cdots)(\mathbf{H}^T\mathbf{H})^{-1} \\
&\approx (\mathbf{H}^T\mathbf{H})^{-1}.
\end{aligned}$$

In the derivation of Eq.(6), we use the following definition of the matrix L2-norm [10].

$$||\mathbf{A}^{-1}|| = 1/\mu_{min}[\mathbf{A}],$$

where $\mu_{min}[\mathbf{A}]$ is the minimum singular values of matrix $\mathbf{A}$,

The regularization parameter δ has the effect of increasing the minimum singular value and reducing the norm of the inverse filter, and this is believed to reduce the sensitivity to the RTF variations. It should be noted that the regularization parameter also reduces accuracy of the inverse filter, and a compromise should be adopted.

2.2.2 *Filter length*

The norm of the filter expressed in Eq. (4) depends on the given filter length M. By increasing the filter length, we can expect to find a filter with the smallest norm among all possible filters.

2.2.3 *Modeling delay*

When a modeling delay d ($d \geq 1$) is used, we also expect the filter norm to be reduced because the causality constraint is relaxed. The filter may correspond to the minimum-norm solution that could be obtained in the frequency domain. Note that, with a frequency domain calculation, sufficiently large number of FFT points should be used to avoid errors induced by the influence of circular convolution.

3. EXPERIMENT

Simulations are made to investigate the effect of the impulse response fluctuations caused by the source position changes.

3.1 Experimental setup

Figure 1 shows the arrangement of the source and the microphones used in the experiment. Room impulse responses between the source and the microphones are simulated by using the image method [11]. The impulse responses are truncated to 1600 samples, corresponding to -60 dB attenuation ($J = 1599$). The sampling frequency is set at 8 kHz, then the duration of the impulse responses is 200 msec. Figure 2 shows an example of the impulse response and its frequency response.

We assume that the source position moves to a new position on the horizontal plane. We refer to the center position before movement as the “reference position”, and refer to the new position after movement as “new position” hereafter. As shown in Fig. 1, equally spaced six new positions are selected that are placed on the circle with radius r centered at the reference position. In [4], it is concluded that small changes in the source position of just a few tenths of the wavelength under consideration can cause large degradations in the equalized response. In this study, the distance from the reference position to the new position, r, is set at 2 cm. This roughly corresponds to one-eighth of the wavelength of the center frequency considered in the simulation. Let the wavelength of the center frequency be λ_c, then,

$$\lambda_c/8 = c/8f_c = 340/(8 \times 2000) = 0.021,$$

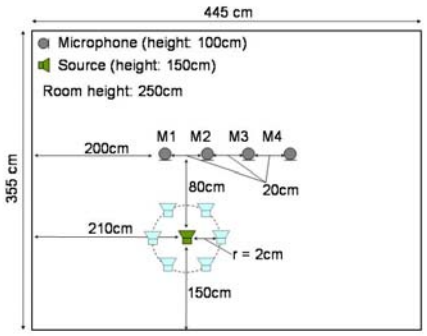

Figure 1: Arrangement of the source and the microphones. M1, M2, M3 and M4 denote the microphones.

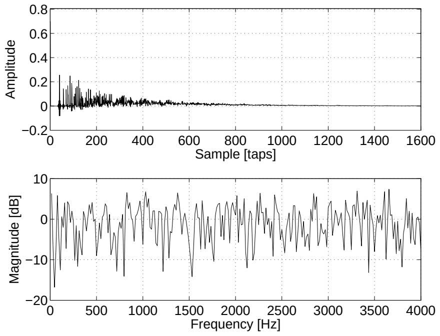

Figure 2: Waveform of an impulse response and its frequency characteristics.

where c is the speed of sound, and f_c is the center frequency.

3.2 RTF variations by position difference

Variations in the room impulse responses caused by changes in the source position are evaluated by using correlation coefficients. The correlation coefficients are calculated between the reference impulse response and the new impulse responses. Here, the reference impulse response represents the impulse response between the reference source position and the microphones, and the new impulse responses represent the impulse responses between the new source positions and the microphones. Then, the correlations are averaged over six positions for each microphone. Figure 3 shows this correlation calculated by using the truncated impulse responses $h_i(n), n_t \leq n \leq J$, where n_t is the truncation point as shown on the time axis. The correlation reaches 0.95 for the truncation point of 300 to 400 taps (40 to 50 ms). The correlation coefficients for all the microphones become large as the analysis period approaches the latter part of the impulse response, and saturates around 700 taps. We can observe that the late reflection part has lower variations as compared with the early reflection part.

3.3 Estimation accuracy of reference RTF

In this simulation, a speech signal with a 6-second duration is used. Reverberant speech signals are simulated by convolving the original speech with the reference impulse responses. The initial 3-second parts of reverberant speech signals are used to estimate the room impulse responses. Speech signals taken from the microphone pairs (1, 2) and (3, 4) are individually used for estimation. Table 1 shows the accuracy of the estimated impulse responses evaluated by using the signal-to-distortion ratio, SDR_{IR}, defined as

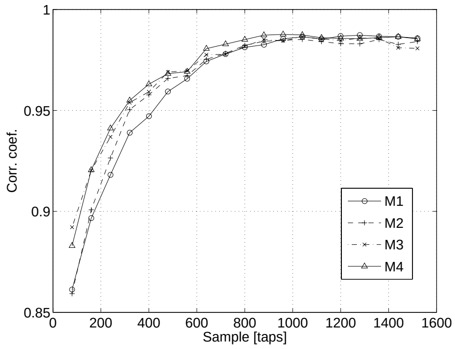

Figure 3: Correlation coefficients as a measure of variations caused by position changes. ’M1’, ’M2’ , ’M3’ and ’M4’ denote the correlations for each microphones.

$$SDR_{IR} = 10\log_{10}\left(\frac{\sum_{n=0}^{J} h^2(n)}{\sum_{n=0}^{J}(h(n)-\hat{h}(n))^2}\right), \qquad (7)$$

where $h(n)$ and $\hat{h}(n)$ are the original and estimated impulse responses, respectively.

Table 1: Estimation accuracy of the impulse responses.

Mic.	1	2	3	4
SDR_{IR} [dB]	49.39	48.78	69.32	68.66

3.4 Evaluation procedure

Dereverberation performance for changes in the source position is evaluated as follows.

(1) An inverse filter set is calculated based on the estimated impulse responses obtained using the reference source position.

(2) For each new source position $j (j = 1, \cdots, 6)$ and each microphone, a reverberant speech signal is simulated by convolving the original speech with the j-th new impulse response.

(3) Dereverberated speech for the j-th position is calculated by filtering the reverberant speech obtained in (2) with the inverse filter set calculated in (1).

(4) SDR values are calculated for all of the dereverberated speech obtained in (3), and these are averaged over six positions to obtain the overall performance measure. This performance measure, denoted as SDR_{SP}, is defined as

$$SDR_{SP} = \frac{1}{6}\sum_{j=1}^{6}\left(10\log_{10}\left(\frac{\sum_{n=0}^{J} s^2(n)}{\sum_{n=0}^{J}(s(n)-\hat{s}_j(n))^2}\right)\right),$$

where $s(n)$ and $\hat{s}_j(n)$ are the original and the dereverberated speech signals for the j-th source position, respectively.

3.5 Effects of design parameters on performance

The reference values for the three design parameters are set at $d = 0$, $M = J$, and $\delta = 0$, respectively. These parameters are individually varied from the reference values to investigate their influences on the performance.

In order to show the effectiveness of the inverse filters when the impulse responses do not change, reverberant speech signals for the reference source position are processed by the inverse filter set designed for the reference source position, and we confirmed that highly precise dereverberation performance is achieved.

Figure 4 shows the effect of modeling delay on the performance and the filter norm. In the non-delay case ($d = 0$), the filter norm for microphone pair (1, 2) becomes very large. This may be because the room transfer functions have maximum phase common zeros on the z-plane. This comes from the symmetrical sound field caused by the positions of the source, M1, M2, and the geometry of the walls, the floor and the ceiling. In such a case, the filter norm is effectively reduced by introducing the proper modeling delay, and the performance improves. This is because the effect of the common zeros is efficiently compensated for by the non-causal part of the filter. When an inverse filter with a relatively small norm has already been obtained as with the (2, 3) and (3, 4) pairs, the influence of this parameter is small. The performance does not change greatly over a large range for the (2, 3) pair, and the performance decreases for the (3, 4) pair. Based on this result, the modeling delay is expected to be effective when the source and microphone geometry is not suitable for inverse filtering.

Figure 5 shows the effect of the filter length on the performance and the filter norm. For the (1, 2) pair, lengthening the filter also reduces the filter norm and improves the performance, although this improvement is not large. The average improvement obtained by lengthening the filter by 500 taps is 2.4 dB. It should be noted, however, that the performance does not decrease with the filter length. It is expected that a longer filter will not have a detrimental effect on the performance.

Figure 6 shows the effect of the regularization parameter on the performance, and the corresponding filter norm. Regularization parameter δ is set at $10^{-1}, \cdots, 10^{-10}$ in calculating the inverse filter. As for the microphone pair (1, 2), the filter norm can be reduced by increasing the regularization parameter, and the perforance also improves. We also observes a small increase in SDR for the microphone pairs (2, 3) and (3, 4). When an overlarge parameter value is used (such as 10^{-1}), however, the performance starts to decrease because the inverse filter becomes innacurate. The adequate value for the microphone pair (1, 2) is $\delta = 10^{-2}$, and is $\delta = 10^{-3}$ for the pairs (2, 3) and (3, 4).

From the experimental results shown above, three design parameters are shown to be effective in making the filter less sensitive to RTF variations.

3.6 Discussion

The above results are obtained when two of the three parameters are fixed to the reference values, namely $d = 0$, $M = J$, and $\delta = 0$. In order to search for more appropriate parameter combination, the influence of the filter length and regularization on the performance is investigated with a modeling delay $d = 200$. Figure 7 shows the performance with various filter lengths and the regularization parameter fixed at $\delta = 0$. The effect of lengthening the filter is observed, and the average improvement realized by lengthening the filter by 500 taps is 4.2 dB. This improvement is greater than that in Fig. 5 (non-delay case), where a 2.4 dB improvement is achieved.

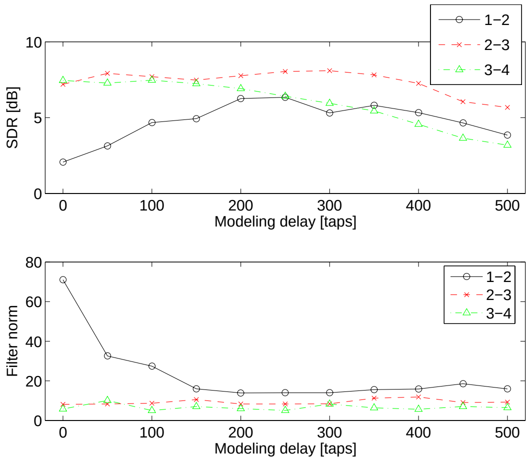

Figure 4: Effect of modeling delay on performance (above), and corresponding filter norm (below).

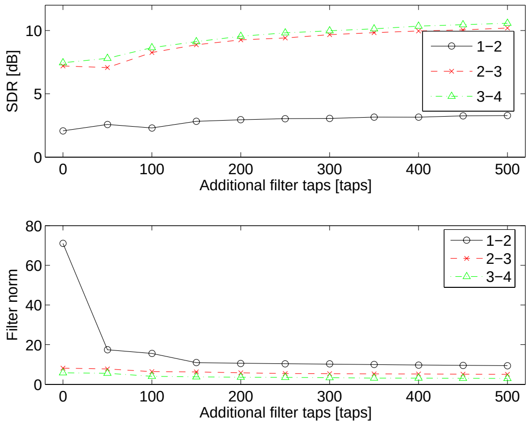

Figure 5: Effect of filter length on performance (above), and corresponding filter norm (below).

Figure 8 shows the performance when regularization is also applied to the above result (filter length: $M = J + 500$, modeling delay $d = 200$). The performance shows similar trends as those in Fig. 6, and reaches its maximum at around 10^{-3}. By properly choosing the regularization parameter, we obtain an average increase of 2.4 dB for all the microphone pairs as compared with $\delta = 0$. For microphone pair (3, 4), SDR reaches 15 dB. From Figs. 6 and 8, we can expect the regularization parameter to be effective especially when the filter is sensitive. We use this adequately designed filter to evaluate the performance for the reference source position (without RTF variation), and obtain an average performance of 25.1 dB, which is around 10 dB higher than the performance with RTF variation.

We also conduct a simulation under the above conditions using the four microphones simultaneously. In this case, the inverse filters are calculated using the four impulse response estimates, and 18 dB is achieved. This result is 3 dB higher

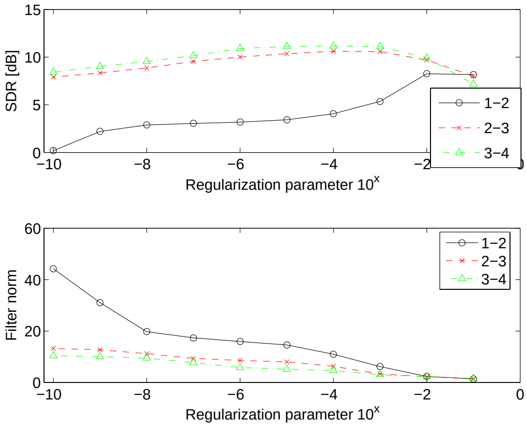

Figure 6: Effect of regularization parameter on performance (above), and corresponding filter norm (below). '-10', ... , '-1' correspond to $\delta = 10^{-10}$, ... , 10^{-1}.

than the performance using the microphone pair (3, 4).

When the RTFs show random fluctuations and its variance is used as the regularization parameter, the filter shown in Eq. (4) gives the optimum solution. Then, the variance of the impulse responses in terms of the position changes is evaluated as follows.

$$Var(n_t) = \frac{1}{6}\sum_{j=1}^{6}\left(\sum_{n=n_t}^{J}(h_c(n) - h_j(n))^2\right), \qquad (8)$$

where $h_c(n)$ and $h_j(n)$ are the reference and the j-th new impulse responses, respectively, and n_t is the truncation point. This variance decreases with increasing n_t, and reaches 10^{-3} around $n_t = 640$ (80 ms). This result implies that the fluctuation in the latter 120-ms part of the impulse response may be regarded as random noise, and that this deviation is mitigated with the regularization parameter.

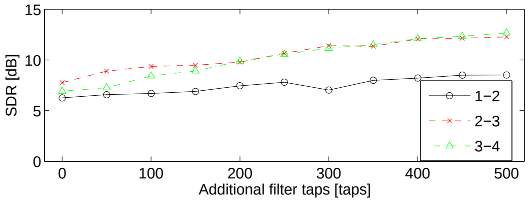

Figure 7: Effect of filter length on performance ($d = 200$, $\delta = 0$).

In our study, room impulse responses are simulated by using the image method [11]. These impulse responses may be different from actual ones. For example, simulated absorption characteristics are flat but actual ones are not. However, we expect that the proposed method may also be applicable to actual impulse responses, and we are planning to conduct experiments in this case. We will also consider the differences between our artificial fluctuations and actual ones observed, for example, during human speech recordings.

4. SUMMARY

In order to extend the applicability of inverse-filter-based dereverberation, this study investigated the effect of the inverse filter design parameters on performance. The filter length, modeling delay, and regularization parameter were arranged to improve the performance when the source position changed. Simulation results showed that the proper choice of design parameters improved the dereverberation performance when a small variation occurred in the RTFs after they had been estimated. The dereverberation performance exceeded 15 dB when the source position changed with one-eighth wavelength distance, where conventional investigations claimed that such a variation would cause a large degradation.

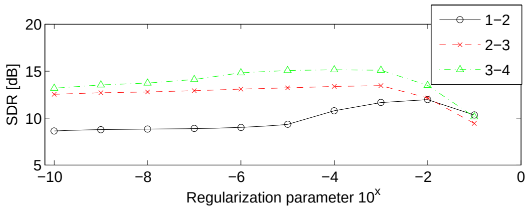

Figure 8: Effect of regularization on performance ($M = J + 500$, $d = 200$). '-10', ... , '-1' correspond to $\delta = 10^{-10}$, ... , 10^{-1}.

REFERENCES

[1] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J. Sound and Vibration*, vol. 102, no. 2, pp. 217–228, 1985.

[2] T. Hikichi and F. Itakura, "Time variation of room acoustic transfer functions and its effects on a multi-microphone dereverberation approach," *Workshop on Microphone Arrays: Theory, Design & Application*, 1994.

[3] M. Omura, M. Yada, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Compensating of room acoustic transfer functions affected by change of room temperature," *Proceedings of the ICASSP IEEE*, vol. 1, pp. 941–944, 1999.

[4] B. Radlovic, R. Williamson, and R. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Trans. SAP*, vol. 8, no. 3, pp. 311–319, 2000.

[5] F. Talantzis and D. Ward, "Robustness of multichannel equalization in an acoustic reverberant environment," *J. Acoust. Soc. Am.*, vol. 114, no. 2, pp. 833–841, 2003.

[6] T. Hikichi, M. Delcroix, and M. Miyoshi, "Speech dereverberation algorithm using transfer function estimates with overestimated order," *Acoust. Sci. and Tech.*, vol. 27, no. 1, pp. 28–35, 2006.

[7] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. ASSP*, vol. 36, no. 2, pp. 145–152, 1988.

[8] M. I. Gurelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. SP*, vol. 43, no. 1, pp. 134–149, 1995.

[9] K. Furuya and Y. Kaneda, "Two-channel blind deconvolution of nonminimum phase FIR systems," *IEICE Trans. Fundamentals*, vol. E80-A, no. 5, pp. 804–808, 1997.

[10] S. Haykin, *Adaptive Filter Theory (Fourth Edition)*. Prentice-Hall, 2002, pp. 436–465.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.