ON-LINE ADAPTIVE ALGORITHM TO ACOUSTIC FLUCTUATION FOR INVERSE FILTER RELAXATION IN SOUND REPRODUCTION SYSTEM

Yosuke Tatekura[†], *Shigefumi Urata*[‡], *Hiroshi Saruwatari*[‡], *and Kiyohiro Shikano*[‡]

[†]Faculty of Engineering, Shizuoka University
3-5-1 Johoku, Hamamatsu, Shizuoka, 432-8561, Japan
email: tytatek@ipc.shizuoka.ac.jp
[‡]Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, 630-0101, Japan

ABSTRACT

This paper proposes a on-line adaptive relaxation algorithm for inverse filter. The inverse filter of the conventional method often amplifies the background noise and line noise when the room transfer function changes. To resolve this problem, an iterative relaxation algorithm for inverse filter by truncated singular value decomposition (Adaptive TSVD) have been proposed. However, it is difficult to apply this relaxation method in a duration where the original signal contains the sound. Therefore, we extend Adaptive TSVD into an on-line type algorithm based on the observed signal at the only one control point, normalizing the observed signal with the original sound. The result of the simulation reveals that the proposed method can always carry out the relaxation process against to acoustic fluctuation in any time durations. Also, subjective evaluation in the real acoustic environment indicates that the sound quality improves without degrading the localization.

1. INTRODUCTION

To achieve a sound reproduction system with loudspeaker reproduction, it is important to design inverse filters which cancel the effects of room transfer functions (RTFs). The RTFs vary depending on environmental variations (such as variations of speed of sound due to temperature fluctuations, change of reflection conditions due to changes in indoor items, etc.) and are not time invariant. Therefore, the reproduction accuracy is deteriorated by environmental variations in a sound reproduction system using a fixed inverse filter coefficients. Also, if unstable inverse filters enlarging the original signal are used, variation of the RTFs cause deterioration of sound quality, it is necessary to either re-estimate the RTFs after variations or to adaptively relax the inverse filters.

As an adaptive design procedure for the inverse filters, a method has been proposed for updating the inverse filter coefficients by means of reference microphones set up at several control points [1]. However, since the reference microphones must be placed near the ears of the listener, hearing is significantly impaired. For re-estimation of the RTFs that vary due to changes in room temperature, a method has been proposed for extending or contracting the time axis of the impulse responses [2]. However, only temperature variations are dealt with and the variation of the reflection due to changes in room configuration cannot be handled.

A method for preventing degradation of sound quality by environmental changes is to relax the inverse filters. In general, inverse filters are designed by deriving the inverse of the matrix consisting of the impulse responses of the RTFs. If the linear independence of the vertical vector comprising this matrix is low, there is a danger that this inverse matrix may expands the solution in the presence of a small amount of error. Methods for resolving this problem include relaxation methods by the regularization method [3] and the truncated singular value decomposition (TSVD) method [4]. In both method, the parameters for performing relaxation of the inverse filters (regularization coefficients or truncation number) can be determined only by considering the matrix of the RTFs. However, it is likely that sufficient control accuracy cannot be obtained because excessively relaxed inverse filters can be designed.

To resolve this problem, we have proposed a method for relaxation of the inverse filters based on adaptive TSVD [5]. In the method, relaxation of the inverse filters is autonomously performed depending on the amount of expansion of the observed noise. Also, the monitoring microphone can be placed at a location that does not restrict the listener. However, it is difficult to adapt the inverse filters in the period which contains speech or music in the original signals. Therefore, we develop the adaptive TSVD into an on-line type algorithm. In this paper, we propose a new on-line adaptive relaxation algorithm for inverse filters based on the observed signal at only one control point, normalizing the observed signal with the original sound.

2. DESIGN METHOD OF RELAXED INVERSE FILTER

2.1 Inverse filter design with Moore-Penrose generalized inverse matrix

In this section, we describe the design method of inverse filter in the frequency domain. In the following, we assume the multi-channel sound reproduction system with M secondary sound sources and N control points.

Let the matrix representing the RTF be $\mathbf{G}(\omega)$, the matrix representing the inverse filter be $\mathbf{H}(\omega)$, the matrix representing the original source be $\mathbf{X}(\omega)$, and the matrix representing the reproduced sound be $\hat{\mathbf{X}}(\omega)$. The matrices can be expressed as follows:

$$\mathbf{G}(\omega)=\begin{bmatrix} G_{11}(\omega) & G_{12}(\omega) & \dots & G_{1M}(\omega) \\ G_{21}(\omega) & G_{22}(\omega) & \dots & G_{2M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ G_{N1}(\omega) & G_{N2}(\omega) & \dots & G_{NM}(\omega) \end{bmatrix} \tag{1}$$

$$\mathbf{H}(\omega)=\begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) & \dots & H_{1N}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) & \dots & H_{2N}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1}(\omega) & H_{M2}(\omega) & \dots & H_{MN}(\omega) \end{bmatrix} \tag{2}$$

$$\mathbf{X}(\omega)=[X_1(\omega),X_2(\omega),\cdots,X_N(\omega)]^T \tag{3}$$

$$\hat{\mathbf{X}}(\omega)=\left[\hat{X}_1(\omega),\hat{X}_2(\omega),\cdots,\hat{X}_N(\omega)\right]^T, \tag{4}$$

where ω denotes the frequency, $G_{ji}(\omega)$ is the RTF and $H_{ij}(\omega)$ is the inverse filter coefficient. $i\ (=1,2,\cdots,M)$ is the order of the secondary sound source and $j\ (=1,2,\cdots,N)$ is the order of the control point. $X_j(\omega)$ is the original sound reproduced at control point j and $\hat{X}_j(\omega)$ is the reproduced sound at control point j.

The reproduced signal $\hat{\mathbf{X}}(\omega)$ can be expressed as follows in terms of the matrices given above:

$$\hat{\mathbf{X}}(\omega)=\mathbf{G}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega). \tag{5}$$

Since our objective is to achieve control such that $\mathbf{X}(\omega) = \hat{\mathbf{X}}(\omega)$ in sound reproduction, the inverse filter $\mathbf{H}(\omega)$ can be derived as the generalized inverse matrix of $\mathbf{G}(\omega)$ in the case of $M > N$. Since the solution becomes underdetermine if there is no rank reduction, the generalized Moore-Penrose (MP) inverse matrix with the least norm solution (LNS) [6] is used. In order to derive the generalized MP inverse matrix, singular value decomposition (SVD) of $\mathbf{G}(\omega)$ is carried out as follows:

$$\mathbf{G}(\omega) = \mathbf{U}(\omega) \cdot \left[\mathbf{\Gamma}_N(\omega), \mathbf{O}_{N,M-N}\right] \cdot \mathbf{V}^{\mathrm{H}}(\omega) \tag{6}$$

$$\mathbf{\Gamma}_N(\omega) \equiv \mathrm{diag}\left[\mu_1(\omega), \ldots, \mu_N(\omega)\right], \tag{7}$$

where $\mathbf{V}^{\mathrm{H}}(\omega)$ is the Hermitian transposed matrix of $\mathbf{V}(\omega)$, $\mathbf{O}_{N,M-N}$ indicates the $N \times (M-N)$ null matrix and $\mu_k(\omega)$ ($k = 1, \ldots, N$, $\mu_k(\omega) \geq \mu_{k+1}(\omega)$) denotes the singular values. By using Eq.(6), the generalized MP inverse matrix of $\mathbf{G}(\omega)$, $\mathbf{G}^{\dagger}(\omega)$, can be given by

$$\mathbf{G}^{\dagger}(\omega) = \mathbf{V}(\omega) \cdot \begin{bmatrix} \mathbf{\Lambda}_N(\omega) \\ \mathbf{O}_{M-N,N} \end{bmatrix} \cdot \mathbf{U}^{\mathrm{H}}(\omega), \tag{8}$$

where

$$\mathbf{\Lambda}_N(\omega) \equiv \mathrm{diag}\left[\xi_1(\omega), \ldots, \xi_N(\omega)\right] \tag{9}$$

$$\xi_k(\omega) = \begin{cases} \frac{1}{\mu_k(\omega)} & \mu_k(\omega) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

By computing the inverse matrix $\mathbf{G}^{\dagger}(\omega)$ for each frequency, the inverse filter $\mathbf{H}(\omega)$ can be designed.

2.2 Relaxation of the inverse filter by TSVD

The inverse filter can be designed by measurement of the impulse responses of the RTFs. However, the measured impulse responses may contain infinitesimal noise. Also, due to environmental variations, the inverse filters may not accurately emulate the inverse characteristics of the actual transfer characteristics. By these reasons, the noise components contained in the original signal may be amplify in sound reproduction. ed in the inverse filter design. Such an amplification of the noise components is not desirable from the viewpoint of sound quality and relaxation of the instability must be performed.

When the generalized MP inverse matrix is derived by SVD, it is necessary to be cautious about the existence of small singular values. This is because such small singular values may contain round-off errors and may have low linear independence, so that the norm of the solution may be expanded. Hence, the solution obtained with all singular values may be unstable. Therefore, by limiting the singular values used in SVD of the matrix, the inverse matrix is stabilized. The inverse matrix with relaxation by TSVD can be obtained by replacing $\xi_k(\omega)$ with 0 even if $\mu_k(\omega) \neq 0$ and $\mu_k(\omega) \cong 0$ in Eq.(10). Then, the number of singular values of the matrix $\mathbf{G}(\omega)$ replaced with 0 is called the truncation number on ω.

3. ON-LINE ADAPTIVE TSVD METHOD

3.1 Outline

In the adaptive TSVD which we have proposed [5], a silent duration in the original sound source is used for the adaptation of the inverse filters. Therefore, this method cannot adapt the inverse filters when the original sound source does not contain the silent duration.

We newly propose an on-line adaptive algorithm which can always adapt in any time duration. Figure 1 shows the overview of the multichannel sound reproduction system of our method. The location of the microphone for signal observation is set apart from the listener not to prevent from listening. In this section, our proposed algorithm, which is called "on-line adaptive TSVD" in this paper, is described. This algorithm introduce the truncation number by the observed signal and relax the inverse filters by TSVD. First, in order to observe the fluctuation of RTFs, silent signal is reproduced at the monitoring microphone. To remove the effect of the original signal, the observed signal is normalized by the original signal. Next, the truncation number of the singular value is obtained by the normalized observed signal, and the relaxed inverse filters are designed.

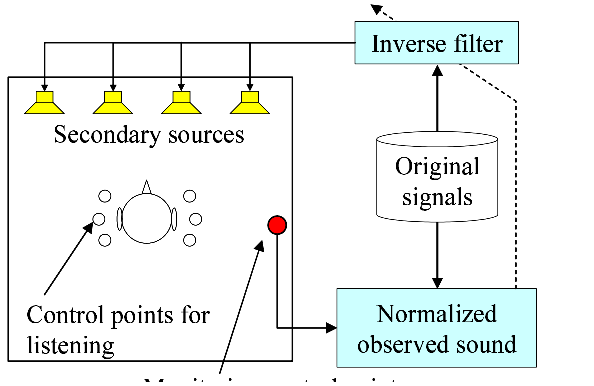

Figure 1: Configuration of sound reproduction system based on the proposed method.

3.2 Reproduction of the silent signal

In case that the RTFs do not fluctuate, the observed signal $\mathbf{X}(\omega)$ can be written as Eq.(5). If the RTFs $\mathbf{G}(\omega)$ are fluctuated into $\hat{\mathbf{G}}(\omega)$, we assume that the fluctuated RTFs $\hat{\mathbf{G}}(\omega)$ is given by

$$\hat{\mathbf{G}}(\omega) = \mathbf{G}(\omega) + \Delta\mathbf{G}(\omega), \tag{11}$$

where

$$\Delta\mathbf{G}(\omega) = \begin{bmatrix} \Delta G_{11}(\omega) & \Delta G_{12}(\omega) & \ldots & \Delta G_{1M}(\omega) \\ \Delta G_{21}(\omega) & \Delta G_{22}(\omega) & \ldots & \Delta G_{2M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Delta G_{N1}(\omega) & \Delta G_{N2}(\omega) & \ldots & \Delta G_{NM}(\omega) \end{bmatrix} \tag{12}$$

as difference between the original RTFs $\mathbf{G}(\omega)$ and the fluctuated RTFs $\hat{\mathbf{G}}(\omega)$. In such case, the observed signal $\hat{\mathbf{D}}(\omega)$ can be expressed as follows:

$$\begin{aligned} \hat{\mathbf{D}}(\omega) &= \hat{\mathbf{G}}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega) \\ &= \mathbf{X}(\omega) + \Delta\mathbf{G}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega). \end{aligned} \tag{13}$$

The observed signal $\hat{D}_N(\omega)$ observed at monitoring microphone is also given by

$$\hat{D}_N(\omega) = X_N(\omega) + \Delta\mathbf{G}_N(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega), \tag{14}$$

where $\Delta\mathbf{G}_N = [\Delta G_{N1}(\omega), \Delta G_{N1}(\omega), \cdots, \Delta G_{NM}(\omega)]$. Here, $\hat{D}_N(\omega)$ is regarded as the error signal obtained by the fluctuated RTFs if $X_N(\omega)$ is zero (silent signal).

3.3 Normalization of the observed signal by the original signals

From Eq.(14), the observed signal $\hat{D}_N(\omega)$ is influenced by the original signals $\mathbf{X}(\omega)$. Because the original signals are time variant, we can hardly obtain only the fluctuation of the RTFs $\Delta\mathbf{G}(\omega)$ from the result of the observed signal. Therefore, we introduce normalized power spectrum of the observed signal by the original signals at i-th iteration, $P^{[i]}_{\text{norm}}(\omega_n)$, which is defined as

$$P^{[i]}_{\text{norm}}(\omega_n) = 10\log_{10}\left\{\frac{1}{L}\sum_{k=n-\lfloor \frac{L}{2}\rfloor}^{n+\lfloor \frac{L}{2}\rfloor} R^{[i]}(\omega_k)\right\}^2 \quad (15)$$

$$R^{[i]}(\omega_n) = \frac{D^{[i]}_N(\omega_n)}{X^{[i]}_{\text{ave}}(\omega_n)} \quad (16)$$

$$X^{[i]}_{\text{ave}}(\omega_n) = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}\left\{X^{[i]}_j(\omega_n)\right\}^2}, \quad (17)$$

where $X^{[i]}_{\text{ave}}(\omega_n)$ is average amplitude spectrum of original signal at control point for listening $X^{[i]}_j(\omega_n)$ $(j = 1,\cdots,N-1)$. Here, L is the window length of the moving average and ω_n is the discrete frequency with index n.

3.4 Update algorithm of inverse filter

Let the truncation number at each stage of iterative updating process be $l_i(\omega_n)$, the variation level of the normalized observed signal be $p_i(\omega_n)$, and the variation magnitude of the truncation number in the stages between $(i-1)$ and i be $a_i(\omega_n)$.
[Step 0] Let us initialize $l_0(\omega_n) = 0$, $p_0(\omega_n) = 1$ and $a_0(\omega_n) = 0$.
[Step 1] An inverse filter $\hat{\mathbf{H}}^{[i]}(\omega_n)$ relaxed by means of $l_i(\omega_n)$ is designed:

$$\hat{\mathbf{H}}^{[i]}(\omega_n) = \mathbf{V}(\omega_n)\cdot\begin{bmatrix}\hat{\mathbf{\Lambda}}^{[i]}_N(\omega_n)\\ \mathbf{O}_{M-N,N}\end{bmatrix}\cdot\mathbf{U}^{\text{H}}(\omega_n), \quad (18)$$

where

$$\hat{\mathbf{\Lambda}}^{[i]}_N(\omega_n) \equiv \text{diag}\left[\frac{1}{\mu_1(\omega_n)},\cdots,\frac{1}{\mu_{N-l_i(\omega_n)}(\omega_n)},\mathbf{o}_{l_i(\omega_n)}(\omega_n)\right]. \quad (19)$$

Here, $\mathbf{o}_{l_i(\omega_n)}$ indicates $l_i(\omega_n)$ zeros. By means of this inverse filter $\hat{\mathbf{H}}^{[i]}(\omega_n)$, the sound is reproduced and the reproduced sound is observed with the monitoring microphone.
[Step 2] The observed signal is normalized using Eq.(15). Also, $p_i(\omega_n)$ is given as

$$\begin{cases} p_i(\omega_n) = +1 & \text{when} \quad T < P^{[i]}_{\text{diff}}(\omega_{n'}) \\ p_i(\omega_n) = -1 & \text{when} \quad P^{[i]}_{\text{diff}}(\omega_{n'}) < -T \\ p_i(\omega_n) = 0 & \text{otherwise} \end{cases}, \quad (20)$$

where

$$P^{[i]}_{\text{diff}}(\omega_n) = P^{[i]}_{\text{norm}}(\omega_n) - P^{[i-1]}_{\text{norm}}(\omega_n). \quad (21)$$

Let n' be

$$n-k \le n' \le n+(W-k), \quad (22)$$

where k is assumed to satisfy $0 \le k \le W$. Also, W expresses the width needed to prevent sharp truncation or untruncation, and T means the threshold for quantization of $p_i(\omega_n)$.

From the result of $p_i(\omega_n)$ and $a_{i-1}(\omega_n)$, $a_i(\omega_n)$ is obtained based on Table 1. So, we can obtain $l_{i+1}(\omega_n)$ as

$$l_{i+1}(\omega_n) = l_i(\omega_n) + a_i(\omega_n). \quad (23)$$

[Step 3] The process is returned to **Step 1** and the original sound is reproduced.

Table 1: Decision rule of $a_i(\omega_n)$

	$a_{i-1}(\omega_n)$		
	-1	0	$+1$
$p_i(\omega_n) = -1$	0	0	$+1$
$p_i(\omega_n) = 0$	0	0	-1
$p_i(\omega_n) = +1$	$+1$	$+1$	-1

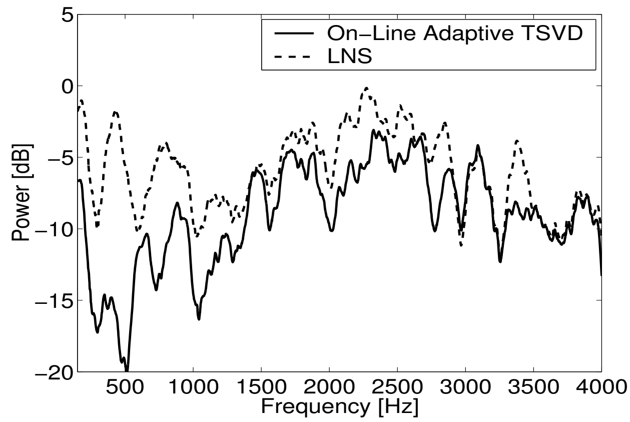

Figure 2: Normalized power spectrum of the observed signal.

4. NUMERICAL EVALUATION

4.1 Experimental setup

To investigate the effect of the proposed method, a numerical simulation with real environmental data was carried out. The system is consisted of 16 secondary sources and 7 control points (6 for listening and 1 for monitoring) within a room with a reverberation time of 0.15 seconds.

In this simulation, the temperature variations in the room are used as the environmental change of the transfer system. The impulse responses measured at a room temperature of 18.0°C is used for the inverse filter design, and those measured at a room temperature of 28.0°C is used as the room impulse response after variations of the transfer system. The inverse filters with 32768 points are designed by the room impulse responses with 9600 points, where sampling frequency is 48000 Hz. The control bandwidth of the inverse filter is 150–4000 Hz. The original signal used for reproduction is a voice signal. The frame length of the observed signal for adaptation is 32768 points. L used in Eq.(15) is 81 points and W in Eq.(22) is 50 points.

4.2 Experimental result

Figure 2 shows the normalized power spectrum of the observed signal before and after adaptation, in which 'LNS' means no adaptation process. From this result, it is found that there is a reduction of the expansion in the low frequency region.

5. SUBJECTIVE EVALUATION

A subjective evaluation test is carried out for comparison of the conventional and the proposed methods in terms of sound quality and sound localization in a sound reproduction system.

The configuration of the sound reproduction system is shown in Fig.3. Six control points for listening are placed at the two ears and 0.05 m in front and behind them, to allow effective sound reproduction even during rotation of the head [7]. The monitoring microphone is placed 0.5 m away from the listener. The 12 primary sources are placed on circle with a radius of 1.5 m. The conditions

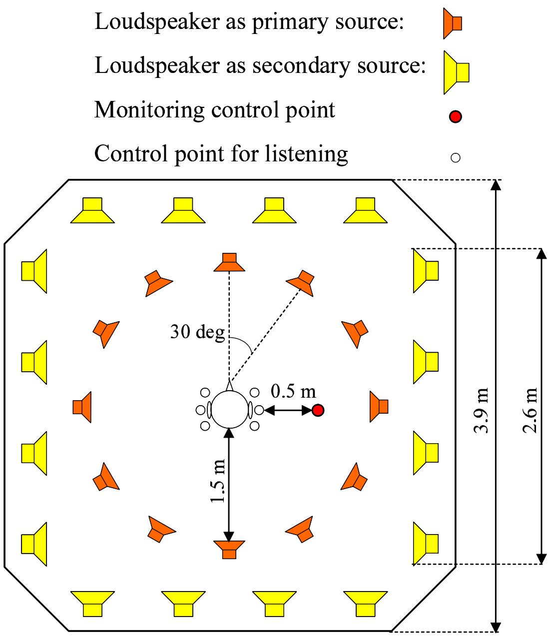

Figure 3: Sound reproduction system for the subjective evaluation.

of the environmental variations are identical to those in the previous section.

To evaluate the effectiveness of applied band in the proposed method, we use five presentation sounds as follows: (1) original signal from the primary sources directly (**Original**), (2) reproduced sound by the inverse filters based on LNS (**LNS**), (3) reproduced sound by the proposed method at the applied band of 150–1000 Hz (**Low**), (4) reproduced sound by the proposed method at the applied band of 1000–4000 Hz (**High**), and (5) reproduced sound by the proposed method at the applied band of 150–4000 Hz (**Full**).

The listeners are 10 males and females with normal hearing capabilities. Evaluation of the sound quality is judged on 5 point evaluation scales (5: very good, 4: good, 3: fair, 2: poor, 1:very poor). With regard to the evaluation of sound localization, any one of 12 directions is judged. First, the listener turns his or her head toward the direction of arrival of the presentation sound. Then, the perceived direction is evaluated.

5.1 Experimental results

The sound quality evaluation results are shown in Fig.4. The error bars in the figure indicate the 95% confidence interval. When analysis of variance is carried out for these results, there is a meaningful difference with a significance level of 5%, but there is no meaningful difference between **LNS** and **High**, and between **Low** and **Full**. This reveals that the sound quality is improved when we apply the proposed method to low band.

Figure 5 shows the correct answer ratios for the perceptual directions. When analysis of variance is carried out for these results, there is no meaningful difference with a significance level of 5% without between **Original** and the others. The result shows that the sound localization of the reproduced sound by the proposed method compares with that by the conventional LNS method.

From these results, we can improve the sound quality of the reproduced sound without degradation of the sound localization.

6. CONCLUSIONS

We described a on-line algorithm for adaptive relaxation of inverse filter by normalizing the observed signal with the original signals. In the simulation using real environmental data, it was found that

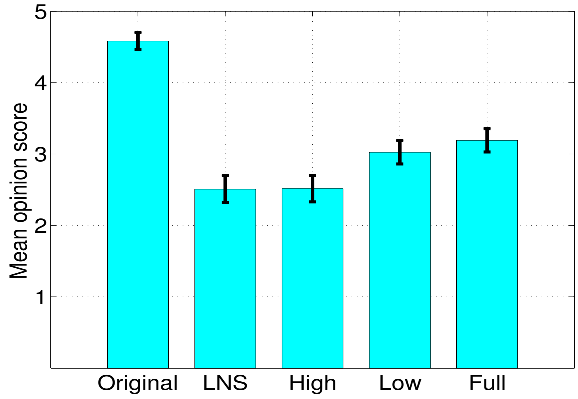

Figure 4: Result of the subjective evaluation for sound quality.

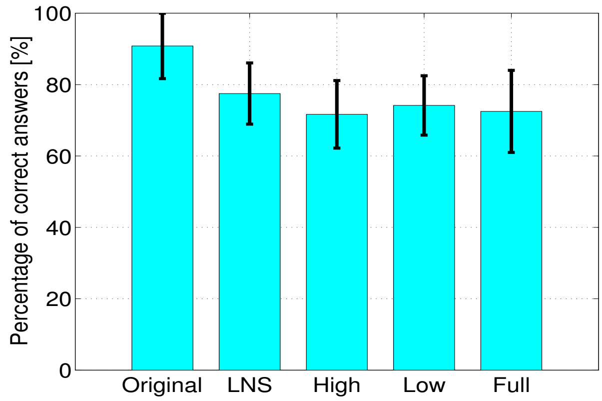

Figure 5: Result of correct answer ratio on perceptual direction.

the proposed method can always carry out adaptation processing to acoustic fluctuation.

REFERENCES

[1] P.A. Nelson, H. Hamada, and S.J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Trans. Signal Processing*, vol.40, no.7, pp.1621–1632, 1992.

[2] Y. Tatekura, H. Saruwatari, and K. Shikano, "Sound reproduction system including adaptive compensation of temperature fluctuation effect for broad-band sound control," *IEICE Trans. Fundamentals*, vol.E85-A, no.8, pp.1851–1860, 2002.

[3] A.N. Tihonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math.*, vol.4, pp.1035–1038, 1963.

[4] P.C. Hansen, "Computation of the singular value expansion," *Computing*, vol.40, pp.185–199, 1988.

[5] Y. Tatekura, Y. Nagata, H. Saruwatari, and K. Shikano, "Adaptive algorithm of iterative inverse filter relaxation to acoustic fluctuation in sound reproduction system," in *Proc. the 18th International Congress on Acoustics (ICA2004)*, Kyoto, Japan, April 2004. (in printing)

[6] A. Kaminuma, S. Ise and K. Shikano, "A method of designing inverse system for multi-channel sound reproduction system using least-norm-solution," in *Proc. Active99*, vol.2, pp.863–874, Florida, USA, Dec. 1999.

[7] A. Kaminuma, S. Ise and K. Shikano, "Robust sound-reproduction-system design against the head movement," in *Proc. WESTPRAC VII*, vol.1, pp.489–492, Kumamoto, Japan, Oct. 2000.