A BROAD-BAND SIGNAL CONTROL IN SOUND REPRODUCTION SYSTEM BY TEMPERATURE COMPENSATION OF ROOM IMPULSE RESPONSES

Yosuke Tatekura, Hiroshi Saruwatari and Kiyohiro Shikano

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, 630-0101, Japan
yosuke-t@is.aist-nara.ac.jp

ABSTRACT
We applied time-warping method [1] to the multichannel sound reproduction system [2] to obtain high-quality sound. In this paper, we apply this method to control a broad-band signal, and evaluate the reproduction accuracy and spectrum of the system response, particularly high-frequency components.

1. INTRODUCTION
It is possible that a sound reproduction system with loudspeaker reproduction could present the desired sound without requiring a special audio device for listeners, unlike headphone reproduction. To achieve such a loudspeaker system, it is important to design multichannel inverse filters which cancel the effects of room transfer functions (RTFs). Generally speaking, in a design of inverse system for such sound reproduction systems, we often consider the RTFs as a time-invariant system. However, the time invariance of the RTFs cannot be guaranteed in a real acoustic environment. Thus, a fluctuation of the impulse responses RTFs degrades the reproduction accuracy of the inverse system. The fluctuation of impulse responses of RTFs is caused by changes in the environment, e.g., the change in temperature or humidity. Temperature fluctuation is the most important factor in the modification of the propagation time of sound because the speed of sound is a function of temperature. The speed of sound is given by

$$c = 331.5 + 0.6T \text{ [m/s]},$$

where $T$ is temperature in degrees Celsius. As shown in Eq. (1), if the room temperature changes by 1 °C, the speed of sound changes by about 0.2%. According to the propagation time of sounds is modified even if transfer channels of sounds do not change; the impulse responses of RTFs also change by about 0.2%. In addition, the exact control of high frequency components is difficult because the wavelength becomes shorter as the frequency increases. For this reason, we should rescale (or "warp") the time axis of the impulse responses of RTFs to compensate the room temperature fluctuation.

2. LINEAR-WARPING PROCESS
Consider the warping of the $N$-point-length impulse response of RTF, $g(n)$, into $g'(m)$ with a warping ratio $C_{th}$. First, $g(n)$ is transformed into a frequency domain by DFT, this can be given by

$$G(k) = \sum_{n=0}^{N-1} g(n)e^{-j\frac{2\pi k n}{N}}, \quad (0 \leq k \leq N-1),$$

where $G(k)$ is a frequency-domain representation of $g(n)$.

Next, we determine $g'(m)$ at every sampled point $m$ ($0 \leq m \leq M - 1$), where $M$ is the maximum integer which does not exceed $C_{th} \cdot N$. If $C_{th} \geq 1$, $g'(m)$ can be given by

$$g'(m) = \frac{1}{N} \left\{ \sum_{k=0}^{N-1} G(k)e^{-j\frac{2\pi k m}{C_{th}N}} + \sum_{k=\frac{N}{2}}^{N-1} G(k)e^{-j\frac{2\pi k (m-N)}{C_{th}N}} \right\}, \quad (3)$$

otherwise $g'(m)$ can be given by

$$g'(m) = \frac{1}{M} \left\{ \sum_{k=0}^{M-1} G(k)e^{-j\frac{2\pi k m}{C_{th}N}} + \sum_{k=\frac{M}{2}}^{M-1} G(k)e^{-j\frac{2\pi k (m-N)}{C_{th}N}} \right\}. \quad (4)$$

The theoretical value of the warping ratio is defined as the ratio of two sound speeds

$$C_{th} = \frac{331.5 + 0.6T_0}{331.5 + 0.6T_t}, \quad (5)$$

where $T_0$ is the original room temperature, and $T_t$ is the room temperature after fluctuation in time $t$.

Hereafter, we regard $g'(m)$ as newly obtained $g(n)$, and its frequency-domain representation, $G(k)$, is used in the next section.
3. INVERSE FILTER DESIGN METHOD

We design inverse filters of the multichannel sound reproduction system in which the number of secondary sources is $I$, the number of control points is $J$. The inverse filters are designed by the least-norm solution (LNS) in the frequency domain [5]. We define the RTF $G(k)$ and the inverse filter $H(k)$ as

$$G(k) = \begin{bmatrix} G_{11}(k) & G_{12}(k) & \cdots & G_{1J}(k) \\ G_{21}(k) & G_{22}(k) & \cdots & G_{2J}(k) \\ \vdots & \vdots & \ddots & \vdots \\ G_{JI}(k) & G_{J2}(k) & \cdots & G_{JJ}(k) \end{bmatrix}$$

$$H(k) = \begin{bmatrix} H_{11}(k) & H_{12}(k) & \cdots & H_{1J}(k) \\ H_{21}(k) & H_{22}(k) & \cdots & H_{2J}(k) \\ \vdots & \vdots & \ddots & \vdots \\ H_{JI}(k) & H_{J2}(k) & \cdots & H_{JJ}(k) \end{bmatrix}$$

where $G_{ji}(k)$ is the RTF between $i$-th secondary source to $j$-th control point, $H_{ij}(k)$ is frequency-domain representation of inverse filter. Here, we can determine $H(k)$ by LNS method and solve

$$H(k) = G^+(k),$$

where $G^+(k)$ is the generalized inverse matrix of $G(k)$. When $I$ is greater than $J$, if the rank of $G(k)$ is $J$, we can determine $H(k)$; this can be given by

$$H(k) = G^H(k) \left\{ G(k) G^H(k) \right\}^{-1},$$

where $G^H(k)$ is the Hermitian transposed matrix of $G(k)$.

4. INVERSE FILTER DESIGN USING REAL ENVIRONMENT DATA

In this section, we design the inverse filter under the assumption that the multichannel sound reproduction system [3], where the number of loudspeakers $I$ is 16 and the number of control points $J$ is 6, using the impulse responses measured in a real acoustic environment.

4.1. Measurement of Impulse Response

The impulse responses used in this study are measured in an acoustic experiment room. The arrangement of apparatus is shown in Fig. 1. Six positions, at the two ears of head and torso simulator (HATS) and 0.05 m in front and behind them, are selected as control points at which the microphones are set, as shown in Fig. 2. Time stretched pulse [4] (TSP) was used as the sound source signal for measurement. The measurement conditions are listed in Table 1. First, we raised the temperature of the room to about 35 °C using an air conditioner. Next, in order to make the room temperature as uniform as possible, the air conditioner was turned off. The impulse responses were measured 20 times at 30 min intervals. The relationship between the times of the temperature observation and observed temperature of each thermometer is shown in Fig. 3, in which T1-T6 denote the thermometers. The temperature converges to a constant at each observed point.

4.2. Linear Warping of Room Impulse Responses

Here, we attempt to warp the impulse responses of RTF, $T_{Ft}$, which were measured at 2.5 hours in those of RTF, $T_{Fo}$, which were measured at 10.0 hours. The temperature room of $T_{Ft}$ is 27.6 °C on average, and that for $T_{Fo}$ is 26.2 °C on average. Although the theoretical value of the warping ratio, $C_{th}$, is calculated to be 1.00240 by Eq. (5), the compensation of RTF based on this value cannot be guaranteed to be suitable because there are bias and errors in the measurement of temperatures. Accordingly we search for suitable warping ratio. Five transfer channels in the sound reproduction system shown in Fig. 1 are chosen, and their impulse responses were warped at a warping ratio ranging from 1.00000 to 1.00400 at intervals of 0.00005. The variations in the SNR between $T_{Fo}$ and warped $T_{Ft}$ are measured. The SNR is defined as

$$\text{SNR [dB]} = 10 \log \frac{\sum_n |d(n)|^2}{\sum_n |d(n) - \hat{d}(n)|^2},$$

where $d(n)$ is the observed signal and $\hat{d}(n)$ is the desired signal. Figure 4 shows the relationship between warping ratios and SNRs at every selected channel, in which a-e represent selected loudspeakers as shown in Fig. 1, and C2 denotes microphone, as

![Fig. 1. Layout of acoustic experiment room.](image1)

![Fig. 2. Arrangement of microphones.](image2)
5. NUMERICAL EVALUATION

5.1. Reproduction Accuracy

In order to investigate the reproduction accuracy of the compensated impulse responses of RTFs, each inverse filter group is convolved with the impulse responses TF0, and the reproduction accuracy for the original source is evaluated using the SNR at each control point. The original sources are obtained by convolving TF0 with INV0, which are ideal responses of this reproduction system.

Table 2 shows the result of the SNRs. In the case of INV0, which is warped by the theoretical warping ratio, the improvement of SNRs is 6.6 dB on the average. In the case of INV0 which is warped by the suitable warping ratio, we can improve the SNR by 14.3 dB by using the compensation method. Also, although the difference between warping ratios is very small, the difference of SNRs is not.

Figure 5 shows the waveform of the system impulse response. These results reveal that the error can be reduced, and in particular, the error after the peak is remarkably decreased. This confirms that the proposed compensation method is effective and indispensable for the realization of an accurate sound reproduction system in a real acoustic environment.

5.2. Spectrum Distortion

The spectral characteristics of the above-mentioned system impulse responses are evaluated. Figure 6 shows the spectral characteristic of system impulse response observed at C2. These results reveal that more error was amplified with higher frequency before using the proposed method, but the spectral characteristics became almost flat after compensation.

To evaluate spectral characteristics numerically, we calculate the spectrum distortion (SD); this can be given by

\[
SD [\text{dB}] = \frac{1}{n} \sum_{k=1}^{n} \left( 20 \log \left| \frac{|D_j(\omega_k)|}{|D_j(\omega_k)|} \right| \right),
\]

where \( |D_j(\omega_k)| \) is magnitude of response of the system obtained by convolving TF0 with INV0 at Cj, and \( |D_j(\omega_k)| \) is that of the system obtained by convolving TF0 with inverse filters, except for INV0.

Table 3 shows the result of the SDs. By using the theoretical warping ratio, we can improve the SDs by about 3.4 dB and by about 3.9 dB by the suitable warping ratio. These results indicate that we must obtain suitable and exact warping ratio to realize high-quality sound reproduction system.
6. CONCLUSIONS

We described a compensation method for temperature fluctuation by linear time warping to avoid the degradation of reproduction accuracy at high frequency components of a sound reproduction system. In numerical simulation using real environmental data, we could improve the reproduction accuracies by about 14 dB using the warping process with a suitable warping ratio, in comparison with those before using the process.

On the other hand, the proposed method requires approximately $N \log N + 4N^2$ multiplications, where $N$ is the length of the compensating RTF in the sample. This heavy computation is a demerit in applying the method to the real-time compensation of the RTF. To introduce the compensation method to an actual sound reproduction system, we must realize a fast compensation algorithm by using, for example, Omura et al.'s method [6].

7. ACKNOWLEDGMENT

This work was partly supported by CREST (Core Research for Evolutional Science and Technology) in Japan.

8. REFERENCES


