Sound Image Reproduction Based on Weighted Room Impulse Responses Using Head-Enclosed Back Surround Loudspeaker Array

Misaki Otsuka¹, Yukoh Wakabayashi², Takahiro Fukumori³, Masato Nakayama⁴ and Takanobu Nishiura⁵

¹ Graduate School of Information Science and Engineering, Ritsumeikan University
² Research Organization of Science and Technology, Ritsumeikan University
³,⁵ College of Information Science and Engineering, Ritsumeikan University
⁴ Ritsumeikan Global Innovation Research Organization, Ritsumeikan University

1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan
E-mail: {¹ is0206pi@ed, ² 17v02111@gst, ³ fukumori@fc, ⁴ mnaka@fc, ⁵ nishiura@is}.ritsumei.ac.jp

Abstract

This study proposes a loudspeaker system called the head-enclosed back surround loudspeaker array (HEBLA) that creates high realistic sound sensations. The proposed system can be combined with video to reproduce a sound field around the listener’s head, but it is difficult to reproduce sound distance cues because the power ratio of direct sound to reverberation at the point of the listener’s ears is different from that in an actual environment. The power ratio is an important cue for perceiving sound distance. Thus, this study proposes a sound image reproduction method which controls sound distance cues by restoring virtual front loudspeakers’ room impulse responses (RIRs) with HEBLA. Evaluation experiments confirmed that the proposed method allows sound distance perception.

1. Introduction

Three-dimensional sound image reproduction systems are used in entertainment systems. Binaural [1] and 5.1-channel surround [2] systems are combined with video to create highly realistic sound sensations. It is important for these systems to faithfully reproduce sound directional and distance cues for accurately perception of a sound source. Sound directional and distance cues allow the listener to locate the position of virtual sound sources. Multi-channel surround systems, such as 5.1-channel surround systems, are commonly used in home theaters. These systems reproduce virtual sound sources in every direction using multiple loudspeakers arranged around the listener, and thus require much space. The authors previously proposed a semi-transaural system called the head-enclosed loudspeaker array (HELA) which reproduces a sound field in a small space. HELA consists of three layers, with loudspeakers three-dimensionally arranged around the listener’s head, as shown in Fig. 1. The previous study [3] used 16 loudspeakers to reproduce a sound field with HELA. In this system, the signal emitted from each loudspeaker is created via the convolution of a dry source and an impulse response from a virtual sound source to each loudspeaker. HELA avoids two problems encountered with binaural systems, namely, headphone discomfort and front-back confusion. The HELA system is not worn on the body, so there is no discomfort. The problem of front-back confusion in sound directional perception is caused by the difference in shapes between a dummy head and the listener’s head. To avoid this problem, HELA takes into account the listener’s head-related impulse response (HRIR), which is an impulse response from the sound source to the listener’s ear. However, HELA is difficult to combine with video because its front loudspeakers obstruct the listener’s view. The present study modifies HELA for use with video. A new loudspeaker system called the head-enclosed back surround loudspeaker array (HEBLA), which lacks front loudspeakers, is proposed. This lack of front loudspeakers makes it difficult to reproduce the sound distance cues created by HELA. Without front loudspeakers, the energy ratio of direct sound to reverberation, an important cue for perceiving the sound distance, is different from that in an actual environment. It is thus necessary to control the energy ratio in HEBLA to reduce this difference. This study proposes a sound image reproduction method for controlling sound distance cues with HEBLA. A conventional sound image reproduction method for HELA is first introduced. Then, a sound image reproduction method for HEBLA is proposed to solve the problems encountered when applying the conventional method to HEBLA. Sound
distance cues are controlled by restoring the virtual front loudspeaker’s information for each loudspeaker in HEBLA. Finally, the effectiveness of the proposed method is evaluated using experiments.

2. Conventional Sound Image Reproduction with HELA

HELA is a semi-transaural system which uses multiple loudspeakers. The signal emitted from each loudspeaker is created via the convolution of a dry source and the room impulse response (RIR) [4] from a virtual sound source to each loudspeaker. RIR includes not only direct sound but also floor and wall reflections and reverberation. Figure 2 shows an example of a virtual sound source being reproduced in front of the listener using the conventional method. The signal $o_e(n)$ heard by the listener is defined by

$$o_e(n) = \sum_{i=1}^{3} \sum_{j=1}^{C} (x(n) * h_{i,j}(n) * k_{e,i,j}(n))$$  (1)

where $C$ is the total number of loudspeakers in the $i$-th layer, $k_{e,i,j}(n)$ is the impulse response from the $j$-th loudspeaker in the $i$-th layer to the listener’s ear, $e \in \{L, R\}$ is the index of the listener’s ear, and $L$ and $R$ are the listener’s left and right ears, respectively. $x(n)$ is the dry source, $n$ is the time index, and $*$ is the convolution operator. $h_{i,j}(n)$ is the impulse response from the virtual sound source to the $j$-th loudspeaker in the $i$-th layer; it is measured in an actual environment or created using the image method [5]. $o_e(n)$ is calculated via the convolution of a dry source and the impulse response from the virtual sound source to the listener’s ear. This method faithfully reproduces the target sound field around the listener’s head by using RIRs. As mentioned above, HELA’s front loudspeakers obstruct the listener’s view. For use with video, a new loudspeaker arrangement is proposed to overcome HELA’s limitations.

3. Proposed Sound Image Reproduction with HEBLA

This section describes the sound image reproduction method for HEBLA. The configuration of HEBLA is presented and some problems are discussed. Then, a sound image reproduction method which controls sound distance cues with HEBLA is described.

3.1 Configuration of HEBLA

This study modifies HELA and proposes a sound image reproduction method for this modified version. The modified system, HEBLA, lacks front loudspeakers. Figure 3 shows the configuration of HEBLA. Although HEBLA can be used with video, it is difficult to reproduce sound distance cues because the power ratio of direct sound to reverberation is different from that in an actual environment. This ratio is an important cue for sound distance perception. The power of direct sound is dominant when a virtual sound source is placed in front of the listener. Thus, removing the front loudspeakers degrades the perception of a virtual sound source in front of the listener. It is possible to construct sound distance cues by controlling the power ratio of direct sound to reverberation in RIRs. This study proposes a sound image reproduction method which controls sound distance cues by restoring the RIR of virtual front loudspeakers. The details of the proposed method are given in the next section.

3.2 Sound image reproduction for front virtual sources

The proposed sound image reproduction method controls sound distance cues for HEBLA by controlling the power ratio of direct sound to reverberation at the point of the listener’s ears by using weighted RIRs. This method weights the original RIR from the virtual sound source to each loudspeaker. The RIR weights are calculated based on the distance between the virtual sound source to each virtual front loudspeaker. Figure 4 shows an overview of the proposed method. The proposed method consists of four processes: determination of RIRs, determination of weights, calculation of weighted RIRs, and creation of output signals for HEBLA. Figure 5 shows an example of sound reproduction in the middle layer. As shown, the proposed method reproduces a virtual sound source in front of the listener with the following steps.
[STEP 1] Determination of RIRs
The RIRs from a virtual sound source to HEBLA are determined using the sound field simulator [3]. The previous study determines the early reflections using the image method [5]. Reverberation is determined using exponentially decaying white noise [3]. In this step, two types of RIR are determined, namely, RIR $h_{i,j}(n)$ from the virtual sound source to the $j$-th loudspeaker in the $i$-th layer and RIR $g_i(n)$ from the virtual sound source to a virtual front loudspeaker. These RIRs are calculated as follows:

$$h_{i,j}(n) = p_{i,j}(n) + r(n - R)$$  \hspace{1cm} (2)
$$g_i(n) = q_i(n) + r(n - R)$$  \hspace{1cm} (3)

where $p_{i,j}(n)$ represents the early reflections determined using the image method, $r(n)$ is the reverberation, $n$ is the time index, and $R (= \text{80 ms})$ is the delay time index. $q_i(n)$ represents the early reflections determined for a virtual front loudspeaker in the $i$-th layer. It is possible to use RIRs measured in an actual environment. However, determining RIRs using the image method allows various conditions to be easily replicated.

[STEP 2] Determination of weights
Weights are determined for each loudspeaker channel. This step is applied to each layer shown in Fig. 3. The loudspeaker channels in each layer are:

- Lower layer: $i = 1, 1 \leq j \leq C$
- Middle layer: $i = 2, 1 \leq j \leq C$
- Upper layer: $i = 3, 1 \leq j \leq C$

where $i (= 1, 2, 3)$ is the layer, $j (= 1, 2, \cdots, C)$ is the loudspeaker channel, and $C (= 7)$ is the total number of loudspeakers in each layer. The arrangement of loudspeakers is shown in Fig. 3. The weights $\alpha_{i,j}$ is calculated by

$$\alpha_{i,j} = \frac{1}{\sum_{j'=1}^{C} 1/d_{i,j'}}$$  \hspace{1cm} (4)

where $d_{i,j}$ is the distance between the $j$-th loudspeaker and the virtual front loudspeaker in the $i$-th layer. Equation (4) shows the ratio of $1/d_{i,j}$ to the sum of $1/d_{i,j'}$ in the target layer. Thus, the weights have a higher value when the $j$-th loudspeaker is close to the front loudspeaker in the same layer. The proposed method is applied to virtual sound sources placed in front of the listener. Since direct sound is a dominant factor for RIRs in the front, the weights emphasize direct sound.

[STEP 3] Calculation of weighted RIRs
Weighted RIRs are calculated for controlling the power ratio of direct sound to reverberation at the point of the listener’s ears. Weighted RIRs are obtained by adding the weighted virtual front loudspeaker’s RIR $g'_i(n)$ to $h_{i,j}(n)$. Weighted RIR $h'_{i,j}(n)$ is calculated as follows:

$$h'_{i,j}(n) = h_{i,j}(n) + g'_i(n)$$  \hspace{1cm} (5)
$$g'_i(n) = \alpha_{i,j} \cdot g_i(n - N)$$  \hspace{1cm} (6)

where $N$ is the delay time index, which is experimentally calculated from the arrival time difference of direct sound. The RIRs from the virtual sound source to the virtual front loudspeakers are restored with other RIRs in HEBLA.

[STEP 4] Creation of output signals for HEBLA
The output signals for HEBLA are created via the convolution of a dry source and each weighted RIR. The output signal $y_{i,j}(n)$ for the $j$-th loudspeaker in the $i$-th layer is calculated by

$$y_{i,j}(n) \equiv x(n) * h'_{i,j}(n)$$  \hspace{1cm} (7)

where $x(n)$ is a dry source. The signal $o_c(n)$ heard by the listener is defined by

$$o_c(n) \equiv \sum_{i=1}^{3} \sum_{j=1}^{C} \left( x(n) * h'_{i,j}(n) * k_{c,i,j}(n) \right)$$  \hspace{1cm} (8)

This shows the impulse response from the virtual sound source to the listener’s ears when the dry source is an impulse signal.

The above steps control the power ratio of direct sound to reverberation at the point of the listener’s ears, allowing the listener to accurately perceive sound distance with HEBLA.

4. Evaluation Experiments
Two evaluation experiments were carried out to confirm the effectiveness of the proposed method. In the experiments, sound distance perception and sound directional perception were evaluated via a comparison with an actual environment.

4.1 Experimental conditions
For each experiment, RIRs were determined using the image method. Table 1 shows the parameters used for RIR determination, and Fig. 6 shows the experimental setup. The reflection order and reflection coefficient were experimentally...
Table 1: Conditions for RIR determination

| Condition            | Value         |
|----------------------|---------------|
| Sound velocity       | 340 m/s       |
| Direction            | 0 deg. (front direction) |
| Distances (d)        | 0.5, 1.0, 1.5, and 2.0 m |
| Reverberation time   | $T_{[60]} = 650$ ms |

Table 2: Conditions for subjective evaluation experiment

| Condition            | Value         |
|----------------------|---------------|
| Sound source         | Japanese female voice (/ikioi/) |
| Number of subjects   | 7 (3 females and 4 males) |

---

determined. The height of a virtual sound source was set to the height of the middle layer of HEBLA. Sound distance perception with the conventional sound image reproduction method [3] was compared to that with the proposed method. For each method, the loudspeaker arrangement of HEBLA, excluding rear-loudspeakers, was used. In the objective evaluation experiment, sound distance perception was evaluated in terms of the direct-to-reverberant energy ratio (DRR) [6], indicates the power ratio of direct sound to reverberation in an impulse response. DRR value was calculated by

$$
\text{DRR} = 10 \log_{10} \frac{\sum_{n=0}^{T_d-1} o^2_e(n)}{\sum_{n=T_d}^{T'-1} o^2_e(n)}
$$

where $o_e(n)$ is the evaluated impulse response, $T_d(=7$ ms) is the length of direct sound, and $T'$ is the length of $o_e(n)$. A higher DRR value indicates that the virtual sound source was perceived close to the listener. DRR values were calculated for the signals heard by the listener, given in Eqs. (1) and (8). In this experiment, $x(n)$ and $k_{e,i,j}(n)$ in Eqs. (1) and (8) were set to an impulse signal and an impulse response from the $j$-th loudspeaker in HEBLA to the dummy head’s left ear (NEUMANN, KU100), respectively. In this experiment, we used DRR value of binaural room impulse response (BRIR). BRIR is an impulse response from a virtual sound source to the dummy head’s ear, and is measured in an actual environment. The impulse response shows same distance perception in case that listeners perceive the virtual sound source in an actual environment. Having the same DRR value with BRIR indicates that the evaluated condition reproduced the sound distance cues of an actual environment.

In the subjective evaluation experiment, sound directional perception and sound distance perception were evaluated. Subjects were asked to listen to an evaluation sound while keeping their eyes closed. They were then asked to give the position they perceived the sound to originate at. Eight sound directions were used (0, 45, ..., 315 deg.). Subjects marked the sound distance and direction on their answer sheet. The results for sound directional perception were evaluated in terms of the correct answer rate. The results for sound distance perception were obtained by calculating the average error distance $D_m$ between the presented distance and the distance given by the listener, defined by

$$
D_m = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} |a_m - b_{s,t}|
$$

where $a_m$ is the presented distance, $m(=0.5, 1.0, 1.5,$ and $2.0$ m) is the sound distance, $S$ and $T(=3)$ are the numbers of subjects and trials, respectively, and $b_{s,t}$ is the sound distance given by the subjects. Table 2 shows the conditions used for the subjective evaluation experiment. The virtual sound sources were reproduced under the conditions given in Table 1. The goal of the proposed method is to reproduce the sound distance cues perceived in an actual environment. The experiment in the actual environment was carried out under the conditions given in Table 1. The HEBLA experiments were carried out in a soundproof room ($T_{[60]} = 150$ ms).

4.2 Experimental results

Figure 7 shows DRR values for each distance. In the figure, the horizontal axis shows the distance between the virtual sound source and the listener, and the vertical axis shows DRR value. When the conventional method was used with HEBLA, DRR values were lower than BRIR values. When the proposed method was used with HEBLA, DRR values were similar to BRIR values, indicating that the proposed
method reproduces the power ratio of direct sound to reverberation. The proposed method is particularly effective for a virtual sound source near the listener. DRR values were higher than BRIR values when the sound distance was more than 1.5 m. As the virtual sound source gets farther away, the power ratio of direct sound to reverberation approaches unity. Direct sound is dominant in the virtual front loudspeaker’s RIR, and thus the proposed method produces DRR values that are higher than BRIR values.

Figure 8 shows the results of sound distance perception. In the figure, the horizontal axis shows the distance between the virtual sound source and the listener, the vertical axis shows the average error distance calculated using Eq. (10), and the error bars indicates the standard deviations. The proposed method reduces the average error distance compared to that for the conventional method for virtual sound sources closer than 1.0 m. This confirms that the weighting process of the proposed method controls sound distance cues. For a sound source distance of 0.5 m, there is a significant difference between the conventional method and proposed method, as determined using the t-test at the 10 % level. The objective evaluation confirms that DRR value is accurately reproduced with the weighting process for virtual sound sources closer than 1.0 m. Sound distance perception is thus improved by controlling the power ratio of direct sound to reverberation.

Table 3 shows the percentage of correct answers for sound directional perception. The scores for the conventional and proposed methods are lower than that for the actual environment. The discrepancy in accuracy is caused by displacement of the listener’s head.

5. Conclusion

This study proposed a sound system called HEBLA and a corresponding sound image reproduction method. The proposed method controls the sound distance cues around the listener’s head by restoring RIRs from a virtual sound source to virtual front loudspeakers. Evaluation experiments confirmed that the proposed method is effective for virtual sound sources near the listener. In the future, we will develop a sound image reproduction method for virtual sound sources far from the listener and in other directions.

Acknowledgment

This work was partly supported by Ritsumeikan University R-GIRO, COI, and JSPS KAKENHI Grant Numbers JP26280065, JP15K16030, and JP16K16094.

References

[1] C.P. Brown and R.O. Duda: A structural model for binaural sound synthesis, IEEE Transactions on Speech and Audio Processing, Vol. 6, No. 5, pp. 476–488, 1998.
[2] G.A. Soulodre, M.C. Lavoie and S.G. Norcross: Objective measures of listener envelopment in multichannel surround systems, 114th Convention of the Audio Engineering Society, Vol. 51, pp. 826-840, 2003.
[3] T. Yoshimura, Y. Wakabayashi, T. Fukumori, M. Nakayama and T. Nishiura: Sound localization with distance perception based on simulated room impulse response using 16ch head-enclosed loudspeaker-array, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing 2017, pp.421–424, 2017.
[4] M. Jeub, M. Schafer and P. Vary: A binaural room impulse response database for the evaluation of dereverberation algorithms, 16th International Conference on Digital Signal Processing, pp. 1–5, 2009.
[5] J.B. Allen and D.A. Berkley: Image method for efficiently simulating small-room acoustics, The Journal of the Acoustical Society of America, Vol. 65, No. 4, pp. 943–950, 1979.
[6] Y.C. Lu and M. Cooke: Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 7, pp. 1793-1805, 2010.