Effect of reducing data amount by reducing quantization resolution of head-related impulse responses on sound localization

Kanji Watanabe*, Shouichi Takane† and Koji Abe‡

Faculty of Systems Science and Technology, Akita Prefectural University,
84-4 Ebinokuchi, Tsuchiya, Yuri-Honjo, 015-0055 Japan

(Received 1 September 2017, Accepted for publication 15 October 2017)

Keywords: HRTF, HRIR, Data reduction, Localization

PACS number: 43.60.-c, 43.66.-x [doi:10.1250/ast.39.316]

1. Introduction

The human auditory system can localize sound sources by using various physical cues, such as the interaural time/level difference (ITD/ILD) and frequency spectrum. Head-related transfer functions (HRTFs) are known to comprehensively include these localization cues [1]. Virtual auditory display systems that present virtual sound sources to listeners can be implemented by the convolution of sound signals with the HRTFs corresponding to the positions of sound sources [2]. Since the HRTFs vary as a function of source direction and distance, the required amount of computational memory increases with the number of virtual sound sources. If the HRTFs could be represented by fewer parameters, the amount of required memory would be greatly reduced. Various investigations using simple representations of HRTFs have been conducted. Langendijk and Bronkhorst [3] removed cues from certain HRTF frequency bands by flattening the magnitude characteristics of the bands. Watanabe *et al.* [4] used a similar approach but the flattening was applied only to contralateral HRTFs. These studies suggested that not all spectral information of HRTFs is necessary for localization. Iida *et al.* [5] represented HRTFs with one peak and two notches and demonstrated via a psychoacoustical experiment that the simplified HRTFs controlled median plane localization well. Although applying the results of these studies will lead to the data reduction of HRTFs, their purposes were mainly to clarify the role of spectral information in sound localization. While the processes in their studies were applied in the frequency domain, there have been few investigations on data reduction processing in the time domain. If the data reduction of HRTFs is effectively implemented by using time-domain processing, it is expected that such processing can be easily implemented in virtual auditory displays (VADs) [6], in which the real-time processing of HRTFs is rigorously demanded.

In this study, the data reduction of HRTFs without influencing localization was investigated more directly. Data reduction was applied to the head-related impulse response (HRIR), which is the representation of the HRTF in the time domain, by reducing the quantization resolution of the instantaneous value of the HRIR. Two subjective evaluations

were conducted using HRIRs with reduced quantization resolution to evaluate the influence on localization and other factors such as timbre.

2. Procedure of HRIR data reduction

Herein, the measured HRIR is denoted as $h(n)$, where h is the instantaneous value for discrete time n . The quantization level of $h(n)$ is defined as 2^x , where x [bit] is the quantization resolution. The HRIR with the reduced resolution, $\hat{h}(n)$, is obtained as follows:

$$\hat{h}(n) = \begin{cases} \left\lceil \frac{h(n)}{2^{A-a}} \right\rceil \cdot 2^{A-a} & (h(n) \geq 0) \\ \left\lfloor \frac{h(n)}{2^{A-a}} \right\rfloor \cdot 2^{A-a} & (h(n) < 0) \end{cases}, \quad (1)$$

where A is the quantization resolution of $h(n)$ and a ($0 \leq a \leq A$) is that for the reduced HRIR. Rounding to the nearest integer in the direction of zero is applied so as not to generate any DC component by this processing. Note that the multiplication by 2^{A-a} is to normalize the maximum value of $\hat{h}(n)$ with that of the original HRIR, $h(n)$. In addition, the operations such as convolution are performed in floating-point arithmetic. A set of HRTFs was measured for each subject using the optimized Aoshima's time-stretched pulse (OATSP) [7] with a sampling frequency of 44.1 kHz and a quantization resolution of 16 bits. In this measurement, the source direction of 0° is defined as the front, and the other directions are defined by clockwise rotation as viewed from above. The source directions had 30° intervals, with 12 azimuths in total. Figure 1 shows the examples of original (quantization level of 16 bits) and data-reduced (quantization levels of 12, 8, 4 bits) HRIRs and their frequency amplitude characteristics. Waveforms are magnified around the rise time of the HRIRs in the left panels. From the figure, the HRIRs with the quantization resolutions of 12 and 8 bits show vary similar characteristics to that with 16 bits. The HRIR with 4 bits shows different characteristics from the original but the outline of the frequency characteristics is retained to some extent.

3. Experiments

To evaluate the influence of data reduction, two subjective evaluations were conducted. One was a localization test, the other was a discrimination test between the stimuli convolved with the original HRIR and the data-reduced HRIR.

*e-mail: kwatanabe@akita-pu.ac.jp

†e-mail: takane@akita-pu.ac.jp

‡e-mail: koji@akita-pu.ac.jp

Fig. 1 Examples of original and data-reduced HRIRs (left panels) and their corresponding frequency amplitude characteristics (right panels) (source direction: 0° , left ear, subject A).

3.1. Common conditions

Five male subjects aged between 21 and 22 years old participated in both of the experiments. They had normal hearing acuity as reported by themselves.

A set of HRTFs (HRIRs) was measured for each subject. The used source signal was pink noise with 2 s duration including 100 ms raised-cosine on- and offset ramps. The stimuli were made by convolving the signal with the original HRIRs or the data-reduced HRIRs. The sound pressure level of the stimulus made by convolving the HRIR at 0° was approximately 70 dB. The level of each stimulus was then adjusted so that the relative level differences among the source directions were maintained. Headphone compensation (equalization) was implemented by dividing the HRTFs into the individual (right and left) headphone transfer functions in the frequency domain. The quantization resolutions were set at 16 (original), 12, 8, and 4 bits.

3.2. Experiment 1: localization test

The number of source directions presented to the subjects was the same as that in the HRTF measurement. Subjects sat in soundproof room and listened to a stimulus presented through headphones (HD-650; Sennheiser Electronic GmbH and Co. KG). The subjects were asked to report the perceived direction in the horizontal plane by writing down a mark on a circle, where the center of the circle corresponded to the subject's head. Each of the stimuli was evaluated 10 times in random order.

As an example, the results of one subject (Subject A) for each quantization resolution are shown in Fig. 2. The abscissa is the simulated direction and the ordinate is the direction perceived by the subject. The quantization resolution is shown at the top of each panel. Despite the increase in front-back

Fig. 2 Scatterplots of presented direction versus localized direction for each quantization resolution.

confusion in some directions (e.g., 210° , 300°), most of the responses lay near the diagonal line.

3.3. Experiment 2: discrimination test

A discrimination test between the stimuli convolved with the original and data-reduced HRIR was conducted to comprehensively evaluate the influence not only on localization but also on other factors (e.g., timbre).

Three stimuli were consecutively presented to the subjects through headphones in the soundproof room. Two of the three stimuli were pink noise convolved with the original HRIR, and the other was the same pink noise convolved with the data-reduced one. The subjects were asked to respond which of the three stimuli was different from the other two in any aspect. A set of stimuli comprising the two original stimuli and a certain stimulus with a reduced quantization resolution at the same source direction was presented in random order. All sets of stimuli were presented five times in random order.

A case in which the subjects discriminated the stimulus convolved with the data-reduced HRIR was regarded as the discriminated answer, and its rate was calculated. Figure 3 shows the average discrimination rates of all subjects for each source direction. For the HRIRs with the quantization resolution of 4 bits, the values are highest. Therefore, the subjects were able to identify the difference caused by the data reduction.

4. Discussion

From the results shown in Fig. 2, the data reduction of HRIRs seems to affect front-back discrimination. On this basis, the front-back confusion rates averaged over all subjects were calculated for all source directions except 90° and 270° . Then, the effects of the quantization resolution and source direction were analyzed by using the analysis of variance (ANOVA). The results showed that neither the main

Fig. 3 Discrimination rate for front (upper panel) and rear (lower panel) source directions.

effects nor the interaction was significant. Therefore, the data reduction of the HRIRs did not influence localization in this experiment. As shown in Fig. 1, the main peaks and notches are retained after reduction even with the quantization resolution of 4 bits. Thus, the subjects are considered to have obtained sufficient localization cues.

In Experiment 2, the subjects chose one out of three stimuli. Therefore, the expected discriminated answer rate is 33%. In Fig. 3, the subjects clearly discriminated the data-reduced HRIRs with the quantization resolution of 4 bits, whereas the discrimination of those with 12 or 8 bits appeared to depend on the source direction. The effects of the quantization resolution and source direction were analyzed by using ANOVA. The results showed that the main effect of quantization resolution was significant ($F(2, 144) = 122.6$, $p < 0.01$). Then, multiple comparison tests by the least significant difference (LSD) method were performed. The results showed that only the quantization resolution of 4 bits differed significantly from the others (significance level was 5%). To determine the causes of the experimental results of Experiment 2, the spectral distortion (SD) between the reference (16-bit) and data-reduced HRTFs was calculated. The SD is defined as:

$$SD = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(20 \log \frac{|H(\omega_k)|}{|\hat{H}(\omega_k)|} \right)^2} \text{ [dB]}, \quad (2)$$

Fig. 4 Spectral distortion for different quantization resolutions.

where N is the sample length of the HRTFs, and $H(\omega_k)$ and $\hat{H}(\omega_k)$ are respectively the frequency responses of the reference (quantization resolution: 16 bit) and data-reduced HRTFs. Figure 4 shows the relationship between the quantization resolution and the mean SD over the subjects and source directions. Note that the extremely large SD values due to the deep notches in $\hat{H}(\omega_k)$ were excluded from the calculation. Such values occurred only when the quantization resolution was 4 bits. The two lines in Fig. 4 indicate the results for left-ear and right-ear HRTFs. As shown in Fig. 4, the change in the SD with the quantization resolution is similar to that in the discrimination rate in Fig. 3, and the SDs for the data-reduced HRIRs with the quantization resolution of 4 bits are much larger than others. Therefore, the stimuli with the quantization resolution of 4 bits were easily discriminated by the subjects.

The two experiments suggest that the data reduction of HRIRs in this study retains the important spectral cues but influences other factors such as timbre when the quantization resolution is too low. In order to minimize the influence of the data reduction, the quantization resolution can be reduced to 8 bits. That is, the data size of original HRIRs obtained at 16 bits may be reduced by half.

5. Conclusion

In this study, we investigated the possibility of the data reduction of HRIRs by reducing the quantization resolution. Two experiments (a localization test and a discrimination test) were conducted using HRIRs with quantization resolutions of 16 (original), 12, 8, and 4 bits. The results suggest the following:

- (1) The data reduction does not significantly influence localization because peaks and notches important for localization are retained.
- (2) For the data reduction not to have any influence as a whole, the quantization resolution may be reduced up to 8 bits.

From the findings of this research, audio data can be compressed by half without deteriorating spatial information. For example, right- and left-channel signals, each of which

having 16-bit quantization, might be assigned to one channel by converting the quantization resolution to 8 bits. This would contribute to realizing multichannel audio systems such as 22.2ch audio [8] with a virtual sound source because data reduction is required in these systems due to the limitation of communication capacity.

An issue that remains for future study is the influence of source signals other than pink noise. Furthermore, an effective representation of HRTF/HRIR utilizing these results is another potential topic for future investigation.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP15K00280.

References

- [1] J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, Mass., 1983), pp. 78–93.
- [2] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic Press Professional, San Diego, 1994), pp. 124–169.
- [3] E. H. A. Langendijk and A. W. Bronkhorst, “Contribution of spectral cues to human sound localization,” *J. Acoust. Soc. Am.*, **112**, 1583–1596 (2002).
- [4] K. Watanabe, R. Kodama, S. Sato, S. Takane and K. Abe, “Influence of flattening contralateral head-related transfer functions upon sound localization performance,” *Acoust. Sci. & Tech.*, **32**, 121–124 (2011).
- [5] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, “Median plane localization using parametric model of the head-related transfer function based on spectral cues,” *Appl. Acoust.*, **68**, 835–850 (2007).
- [6] K. Watanabe, Y. Oikawa, S. Sato, S. Takane and K. Abe, “Development and performance evaluation of virtual auditory display system for dynamically presenting multiple sound images using graphics processing unit,” *J. Acoust. Soc. Jpn. (J)*, **71**, 571–580 (2015) (in Japanese).
- [7] Y. Suzuki, F. Asano, H. Y. Kim and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *J. Acoust. Soc. Am.*, **97**, 1119–1123 (1995).
- [8] Rec. ITU-R BS.2051 (2014).