International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2024)
T. Tanabe and V. Yem (Editors)

Poster

Study of the Optimal Reproduction Method for Living Environment Sources Recorded using Ambisonics

T. Okada [1] and M. Yamataka [1]

[1] Aichi University of Technology, Japan

Abstract

The three-dimensional audio technology called Ambisonics has garnered considerable attention as a high-definition sound space reproduction technique for a sense of presence in virtual reality (VR). In this study, we examined how different sound reproduction methods and source positions affect sound space perception in a headphone-based listening environment. Therefore, environmental sound sources recorded using the Ambisonics method were changed to various reproduction methods using signal processing. The sound sources were converted into four types of reproduction methods: monaural, two stereo reproductions (figure-of-eight and cardioid), and binaural reproduction with Samurai head-related transfer functions (HRTFs). The results showed that the cardioid and binaural playback methods were rated higher in terms of stereoscopic effect and sense of presence compared to the other playback methods. Additionally, it was observed that the cardioid and binaural playback method could achieve greater accuracy in sound localization.

CCS Concepts

• ***Human-centered computing*** → *mixed/augmented reality; Virtual reality;* • ***Hardware*** → *Sound-based input/outputt;*

1. Introduction

In recent years, with the advancement of virtual reality (VR) technology, a 3D spatial audio technology called Ambisonics [Pol05] has garnered considerable attention. This technology analyzes and reproduces plane waves that converge at a listening point by using spherical harmonic transforms, and its playback accuracy depends only on the recording order. This allows the reproduction of high-frequency sound spaces with a system of a certain scale. Consequently, the Ambisonics technology is expected to be applied in various VR environments, including YouTube [SOT*12], [WA01].

However, comparisons with monophonic and stereo playback methods have not been performed in previous studies on the Ambisonics recording and playback methods, and the relationship between the spatial arrangement of sound stimuli during recording and the optimal playback method remains unexplored. In this study, we investigated the effect of different sound reproduction methods and sound source positions on sound space perception when using various reproduction methods, assuming a listening environment using headphones.

2. Generating experimental stimuli

The recording layout is shown in Figure 1. A microphone (AMBEO VR MIC, Senheiser) was placed in front of a coffee shop counter to record environmental sounds, including the sound of coffee pouring, background music, and people's voices. The audio signals recorded using the VR MIC were captured in a four-channel AmbiX format, including the XYZ directional components ($AmbiX_X$, $AmbiX_Y$, $AmbiX_Z$) and an omnidirectional W component ($AmbiX_W$). The recorded sound sources were subsequently cut and edited using the DAW software (REAPER, Cockos), with the stimulus length set to approximately 90 seconds.

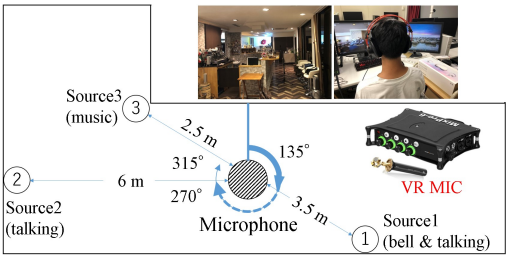

Figure 1: *Recoding layout.*

The playback methods used were monophonic with almost no spatiality (1: Mono), two types of playback methods, i.e., figure-of-eight (2: FoE) and cardioid (3: Car), and a binaural playback method (4: Bin) [Mø92]. Headphones (HDA200, Senheiser) were used to present the experimental stimuli, with the average loudness adjusted to 65±3 dB at the listening point. The corresponding stimuli were created from the recorded AmbiX sound source using the

© 2024 The Authors.
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.2312/egve.20241382

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

following formulae (1)-(4).

$$Mono : AmbiX_W component \tag{1}$$

$$\begin{aligned} FoE : Lch &: AmbiX_X \times \cos 45° + AmbiX_Y \times \sin 45° \\ Rch &: AmbiX_X \times \cos 45° - AmbiX_Y \times \sin 45° \end{aligned} \tag{2}$$

$$\begin{aligned} Car : Lch &: (AmbiX_X \times \cos 45° + AmbiX_Y \times \sin 45° + AmbiX_W)/2 \\ Rch &: (AmbiX_X \times \cos 45° - AmbiX_Y \times \sin 45° + AmbiX_W)/2 \end{aligned} \tag{3}$$

$$Bin : ConvolutedwiththeSamuraiHRTF \text{ [Hir14]} \tag{4}$$

3. Subjective evaluation experiments

Ten university students aged 20–23 years with normal hearing participated in this experiment as listeners. They were asked to evaluate their impressions of the experimental stimuli using headphones. The participants were instructed to rate the following items on a five-point scale ranging from 0 (not at all) to 4 (extremely):

- Stereoscopic effect: The depth and sense of spaciousness, that is, the perception that the sound was coming from afar.
- Sense of presence: The perception that the listener was present when the sounds were generated, including the emotion and satisfaction derived from the sounds.
- Sense of envelopment: The feeling of being surrounded by sound.

Furthermore, the subjects were asked to listen to three types of sounds and answer questions about the direction from which these sounds were coming. They had to choose from eight directions in 45 degree increments relative to the front. The three types of sounds included the following: (1) the sound of the bell at the entrance and speaking position of the person entering, (2) the sounds coming from the positions where two customers were talking, (3) the music being played.

4. Results and discussion

Figure 2 shows the average evaluation scores for the stereoscopic effects. A one-factor analysis of variance (ANOVA) was conducted for each sound source with the four playback methods as the within-subject factors. The result revealed the significant effects of both evaluation for stereoscopy ($F_{3,27} = 4.69$, $p < .01$) and sense of presence ($F_{3,27} = 7.38$, $p < .01$). Multiple comparisons (LSD method, $p < .05$) on the main effect of playback methods revealed that the Car and Bin playback methods significantly higher than the Mono and FoE playback methods in terms of stereoscopy and sense of presence.

Figure 3 shows the angular distribution of the sound localization at the entrance of the coffee shop as a representative example (where the correct answer is 135 degrees). The horizontal axis shows the four playback methods, the vertical axis shows the perceived angles, and the size of each circle indicates the number of listeners. For all the sound sources, the Mono playback methods exhibited a dispersed distribution, suggesting that the direction could not be accurately perceived. In contrast, while using the Car and Bin playback method, the sounds in the left-right direction were not muddled with the sounds in the front-back direction, confirming its relatively high accuracy of directional localization.

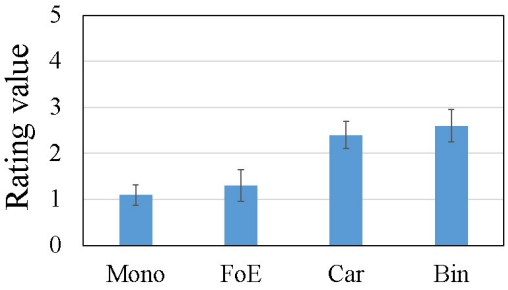

Figure 2: *Results of the stereoscopic effect rating.*

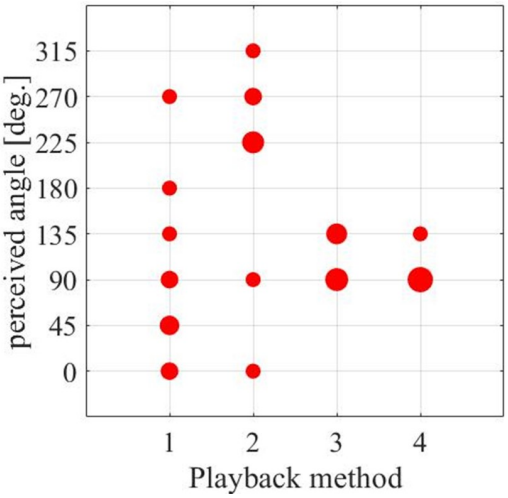

Figure 3: *Results of the sound localization.*

5. Summary

In this study, it was observed that the cardioid and binaural playback methods received higher evaluations in terms of stereoscopic effect and sense of presence. Furthermore, results on directional perception also indicated that the cardioid and binaural playback methods improved sound image localization.

References

[Hir14] HIRAHARA T.: Dynamic binaural playback technology. *The journal of the Institute of Image Information and Television 68*, 8 (2014), 608–611. 2

[Mø92] MØLLER H.: Fundamentals of binaural technology. *Applied Acoustics 36* (1992), 171–218. 1

[Pol05] POLETTI M. A.: Three-dimensional surround systems based on spherical harmonics. *Audio Engineering Society 53*, 11 (2005), 1004–1025. 1

[SOT*12] SUZUKI Y., OKAMOTO T., TREVINO J., CUI Z., IWAYA Y., SAKAMOTO S., OTANI M.: 3d spatial sound systems compatible with human's active listening to realize rich high-level kansei information. *Interdisciplinary Information Sciences Journal 18*, 2 (2012), 71–82. 1

[WA01] WARD D. B., ABHAYAPALA T. D.: Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Trans, Speech and Audio Proc. 9*, 6 (2001), 697–707. 1

© 2024 The Authors.
Proceedings published by Eurographics - The European Association for Computer Graphics.