High-order Ambisonics auditory displays for the scalable presentation of immersive 3D audio-visual contents

Jorge TREVINO*
Research Institute of Electrical Communication, Graduate School of Information Sciences Tohoku University

Takuma OKAMOTO†
National Institute of Information and Communications Technology

Cesar SALVADOR*
Research Institute of Electrical Communication, Graduate School of Information Sciences Tohoku University

Yukio IWAYA‡
Faculty of Engineering Tohoku Gakuin University

Zhenglie CUI*
Research Institute of Electrical Communication, Graduate School of Information Sciences Tohoku University

Shuichi SAKAMOTO*
Research Institute of Electrical Communication, Graduate School of Information Sciences Tohoku University

Yôiti SUZUKI§
Research Institute of Electrical Communication, Graduate School of Information Sciences Tohoku University

ABSTRACT

In recent years, we have conducted research surrounding the harmonic decomposition of sound fields for their encoding and reproduction. As part of this research, we have designed and built high-precision 3D auditory displays using a sound field reproduction technology known as High-order Ambisonics (HOA). An important advantage of HOA over other alternatives is the definition of a system-independent encoding. This allows to fully separate the recording and reproduction stages and, therefore, reproduce the same contents using diverse presentation technologies.

In concrete, we demonstrate two auditory displays based on HOA. One is a large but transportable system consisting of 32 loudspeakers regularly distributed around a listening region. The other one is a compact system using headphones to present the spatial sound information encoded in the Ambisonics format. This is accomplished by the binaural rendering of 32 virtual loudspeakers. Their distribution corresponds to that of the physical array used in the larger system.

Both systems were combined with appropriate video presentation subsystems to allow for the highly immersive presentation of 3D audio-visual contents. In the case of the large, 32-channel system, a sound-transparent screen is used in conjunction with a 3D video projection system using shutter glasses. The compact, headphones-based system uses a head-mounted display to present video.

Both systems are expected to convey a comparable experience when presenting immersive audio-visual contents. To illustrate this, we prepared a set of demonstrations. Three of them feature 3D video, while other two consist of audio-only contents. In all cases HOA was used to render full-surround 3D sound.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities; H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing—Signal Analysis, Synthesis, and Processing

*e-mail: {jorge,salvador,sai,saka}@ais.riec.tohoku.ac.jp
†e-mail: okamoto@nict.go.jp
‡e-mail: iwaya.yukio@tjcc.tohoku-gakuin.ac.jp
§e-mail: yoh@riec.tohoku.ac.jp

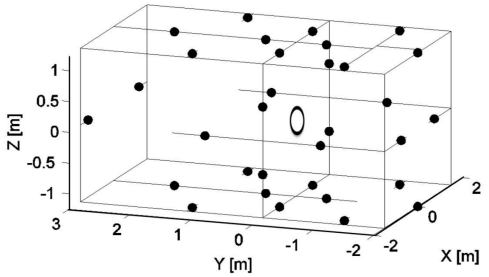

Figure 1: The 32-channel loudspeaker array. Filled circles show the positions of the loudspeakers. The hollow circumference at the center marks the ideal listening position. The screen hangs to the left, on the positive side of the $y-$axis, parallel to the $x-z$ plane.

1 INTRODUCTION

The rapid advancement of computer and multimedia technologies have increased the demands for more realistic presentation systems. End-users seek to experience highly immersive contents and communication systems achieving a very high sense-of-presence. Researchers investigating the nature of human perception require ultra-realistic multimedia systems. The ability to accurately localize objects is a fundamental property of the human visual and auditory systems. Furthermore, both senses must be stimulated in a consistent, precise way to achieve the realisitc presentation of an audio-visual scene.

2 THE LARGE SYSTEM: 32-CHANNEL LOUDSPEAKER ARRAY

We demonstrate a large but transportable system consisting of 32 loudspeakers and a 3D video projection subsystem. Sounds can be presented from all directions around the listener. The loudspeakers are regularly distributed around the listener; their positions are illustrated in Fig 1. They are driven using a technology known as High-Order Ambisonics (HOA). Ambisonics originated as a technique to accurately present sound from all directions using a loudspeaker array [1]. Later, the original formulation evolved into HOA, a sound field reproduction method [2, 3]. Systems based on HOA attempt to re-create the whole sound field, that is, the sound pressure at every spatial position within the listening region. This results in a very natural and realistic presentation of spatial audio. The listener perceive the same sound they would in the real environment, with sounds changing in a physically correct way as the users rotate their heads or change their postures.

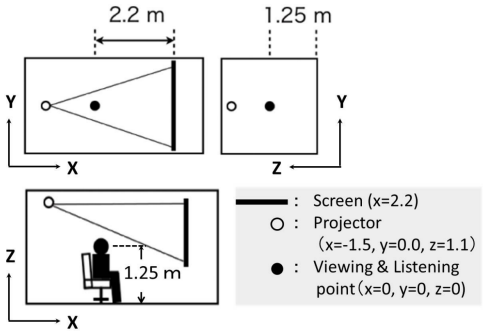

Figure 2: The 3D video projection system. The projector and screen positions were selected to avoid shadowing from the listener when they are inside the listening region.

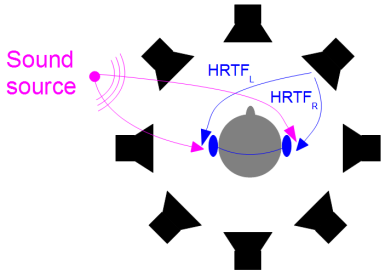

Figure 3: Reproduction of Ambisonics using a virtual loudspeaker array. Each loudspeaker is presented binaurally using the Head-Related Transfer Function measured or calculated at its position. The result of summing all virtual loudspeaker signals is a binaural version of the Ambisonics system; sound sources can be heard from all directions around the user.

2.1 3D video projection

A 3D video projection system using shutter glasses and a sound transparent screen is used to complement the spatial audio contents with video information. The setup for the video system is shown in Fig. 2. A synchronization signal is transmitted through an optical fiber cable when the video starts playing to initiate sound reproduction in the audio subsystem.

The full demonstration system is a transportable version of the 157-channel loudspeaker array and 3D projection system of the Research Institute of Electrical Communication in Tohoku University [4]. The larger system is capable of presenting HOA up to fifth order, a measure of its spatial accuracy. The transportable system can reproduce Ambisonics up to fourth order.

3 The Compact System: Virtual Loudspeaker Array

A virutal version of the 32-channel system was also prepared for headphones listening. This version allows users to experience our demonstration contents even if there is no enough space available to deploy the transportable loudspeaker array. Video contents, supporting both 2D and 3D, are presented using a head-mounted display in place of the projection system.

To present spatial audio contents using headphones, the system creates a virtual loudspeaker array [5]. A schematic of the system is shown in Fig. 3. At its core, it uses what is known as the head-related transfer function (HRTF). These functions characterize sound transmission from a sound source to the listener's two ears. They are used in our system to simulate the transmission path, the presence of the listener's head and body, and the effects of the pinna for each of the 32 loudspeakers of the physical system.

Figure 4: Three of the immersive contents used with our demo system. All of them feature 3D surround sound synthesized using High-Order Ambisonics and 3D video. They can be presented on both, the large loudspeaker array and the virtual presentation system.

Figure 5: A photograph of the demo system. The 32-channel loudspeaker array is distributed over a metal frame housing the screen and listening space. The system is controlled by two computers, one for video presentation and one for sound rendering.

4 Demonstration Contents

To demonstrate the capabilities of our system, we prepared a series of immersive audio samples. Three of them are accompanied by 3D video. A computer simulation of moving sound sources was encoded using fourth-order Ambisonics and prepared for its reproduction using either the large, 32 surrounding loudspeakers systems or the compact headphones-based one. The multimedia contents available include highly localized sources such as falling spheres, as well as ambiental sounds such as forest sounds at night. Objects produce localizable sounds even after they have left the screen, resulting in a highly-immersive experience. Screenshots of these multimedia demos are shown in Fig. 4. Meanwhile, a photograph of the transportable loudspeaker array and projection system, including its control desk is shown in Fig. 5.

Besides these audio-visual contents, two audio-only demonstrations have also been prepared. One of them is an interactive presentation capable of rendering sound from any direction the user chooses. The other one exemplifies sound field reproduction using the recordings of an Ambisonics microphone array.

Acknowledgements

This study was partly supported by the GCOE program (CERIES) of the School of Engineering, Tohoku University, a Grant-in-Aid of JSPS for Specially Promoted Research (no. 19001004) and a Grant-in-Aid of JSPS for Scientic Research (no. 24240016).

References

[1] M.A. Gerzon, *J. Audio Eng. Soc.*, 21(1), 2–10, 1973.
[2] M.A. Poletti, *J. Audio Eng. Soc.*, 53(11), 1004–1025, 2005.
[3] J. Daniel, *Proc. 23rd Int. Conf. Audio Eng. Soc.*, 16, 1–15, 2003.
[4] T. Okamoto, Z. Cui, Y. Iwaya and Y. Suzuki, *Proc. IEEE IC-NIDC*, 179–183, 2010.
[5] M. Noisternig, T. Musil, A. Sontacchi and R. Höldrich, *Proc. 24th Int. Conf. Audio Eng. Soc.*, 1–5, 2003.