IMPROVING SOUND FIELD REPRODUCTION IN A SMALL ROOM BASED ON HIGHER-ORDER AMBISONICS WITH A 157-LOUDSPEARKER ARRAY

T. Okamoto[1,2], *D. Cabrera*[4], *M. Noisternig*[5], *B. FG Katz*[6], *Y. Iwaya*[1,3] *and Y. Suzuki*[1,3]

[1]Research Institute of Electrical Communication, Tohoku University
[2]Graduate School of Engineering, Tohoku University
[3]Graduate School of Information Sciences, Tohoku University
2-1-1, Katahira, Aoba-ku, Sendai, 980-8577, Japan
{okamoto@ais.,iwaya@,yoh@}riec.tohoku.ac.jp

[4]Faculty of Architecture, Design and Planning, The University of Sydney
Sydney, NSW, 2006, Australia
densil.cabrera@sydney.edu.au

[5]IRCAM – CNRS UMR STMS
1, place Igor-Stravinsky, 75004, Paris, France
markus.noisternig@ircam.fr

[6]LIMSI – CNRS
BP 133, 91403, Orsay, Cedex, France
brian.katz@limsi.fr

ABSTRACT

This article presents a case study of higher-order Ambisonics (HOA) for real-time sound field reproduction in a small room with a 157-loudspeaker array. It addresses a number of specific questions and practical issues on the system design and implementation, such as the reproduction room's acoustic, loudspeaker positioning and radiation patterns, distributed computing and audio channel synchronization, and in more general the achievable accuracy of sound field reproduction.

In the current configuration of the system Ambisonics up to order $n = 6$ is applied and the decoders are rendered in parallel on a cluster of four computers. For this reason, synchronization and communication between the different computers becomes a challenging task for achieving a good system performance. The overall system latency and the inter-channel synchronicity have been measured using time-stretched pulse (TSP) signals. The measurement results have shown a maximum (unsigned) latency of 51 samples, which corresponds to $\Delta t = 1.1\,ms$. It is obvious that the acoustic of the reproduction room has a strong effect on the accuracy of the Ambisonics sound field reproduction. To achieve semi-anechoic conditions sound absorption materials have been installed in the room. Finally, spatial filters have been applied to each individual loudspeaker to correct for different orientations with reference to the sweet spot. These filters have been derived from radiation pattern measurements in an anechoic chamber.

1. INTRODUCTION

The authors have developed a 157-loudspeaker array system for sound field reproduction in a small rectangular room with dimensions $x = 5.18\,m$, $y = 3.38\,m$, and $z = 2.52\,m$[1, 2, 3]. All loudspeakers (FE83E; Fostex) are installed on a regular rectangular grid with distance $d_w = 30\,cm$ from the walls and a distance of $\Delta d_{LS} = 30\,cm$ from each other. Fig. 1 shows the surrounding loudspeaker array and Fig. 2 illustrates the arrangement of loudspeakers on the grid. The measured mid-frequency reverberation time RT30 of the untreated reproduction room was about $T_{30} = 0.2\,s$. The audio rendering cluster consists of fourteen 12-channel D/A converters (HD192; MOTU) connected to four computers (Mac Pro; Apple). Clock synchronization is achieved by using a master clock generator (Nanosynsc HD; Rosendahl) connected to each D/A module. Fig. 3 shows a process chart of the developed system.

Several different approaches to sound field reproduction using multi-channel loudspeaker arrays have been studied, such as wave field synthesis [4], boundary surface control [5], and Ambisonics [6]. The reproduction accuracy of wave field synthesis and boundary surface control strongly depends on the spacing between the independently controllable drivers. Applying these methods to the proposed system without any inverse filtering limits accurate sound field reproduction to a frequency band below 350 Hz. However, sound field reproduction with Ambisonics primarily depends on the achievable reproduction order, and therefore on the number of independently driven loudspeakers, and the uniformity of their distribution on the surrounding sphere.

In the following, practical implementation issues including the achievable decoding order for the arrangement of loudspeakers, synchronization of audio channels, influence of the reproduction room's acoustics, and the equalization of the loudspeakers are discussed.

2. APPROPRIATE DECODING ORDER FOR A REGULAR RECTANGULAR GRID ARRANGEMENT

Sound field reproduction with Ambisonics primarily depends on the number of independently driven loudspeakers and the regularity of their angular distribution around the listening area. However, as the loudspeakers of the system are distributed on

Figure 1: Appearance of the surrounding loudspeaker array

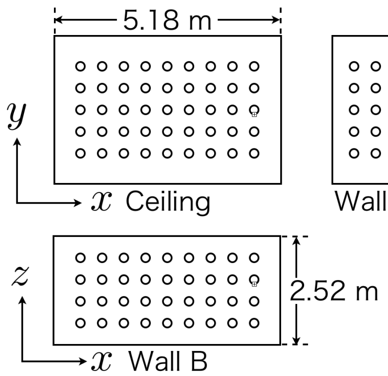

Figure 2: Arrangement of the loudspeakers

a regular rectangular grid, not on a sphere (and with no loudspeakers on the floor), the highest possible reproduction order is limited.

The general Ambisonics approach is defined by a spherical harmonics expansion of the free field without taking the radial solutions of the wave equation – the spherical Bessel or Hankel functions of the first and second kind – into account [7]. The decoding matrix $\mathbf{C}^+$ is calculated from the set of spherical harmonics $Y_n^m(\theta, \phi)$ of order n and degree m for each loudspeaker position applying mode-matching

$$\mathbf{C}^+ = \begin{pmatrix} Y_0^0(\theta_1, \phi_1) & \cdots & Y_0^0(\theta_K, \phi_K) \\ Y_1^{-1}(\theta_1, \phi_1) & \cdots & Y_1^{-1}(\theta_K, \phi_K) \\ \vdots & \ddots & \vdots \\ Y_N^N(\theta_1, \phi_1) & \cdots & Y_N^N(\theta_K, \phi_K) \end{pmatrix}^+ , \quad (1)$$

where K denotes the number of independent loudspeaker channels, N the decoding order, and the operator $(\cdot)^+$ the pseudo-inverse of a matrix.

To determine the maximum decoding order for a given loudspeaker arrangement, the conditioning of the matrix $\mathbf{C}$ is considered as an important factor to estimate the accuracy of decoding. The sweet spots is located in the center of the room at the coordinates $x = 2.59\,m, y = 1.69\,m$ for a seated person's ear height with $z = 1.26\,m$. The condition number depends on the position of the loudspeakers and can be easily determined by projecting the regular grid on a virtual surrounding sphere; and different propagation delays are compensated applying fractional delays. Fig. 4 shows the condition number up to 11th order; due to the

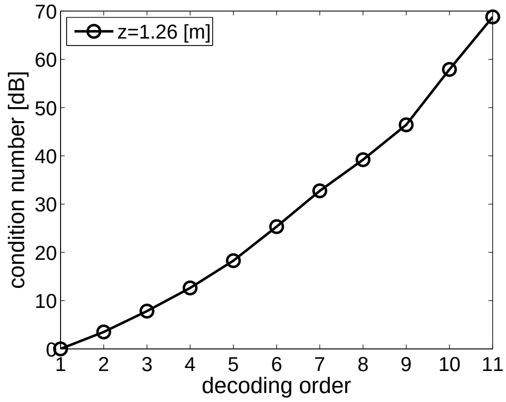

Figure 4: The relationship between condition number of matrix $\mathbf{C}$ and HOA decoding order

non-uniform distribution of loudspeakers and the limitation to the upper hemisphere, the condition number is relatively high even for lower reproduction orders. Optimal decoding for hemispherical loudspeaker arrangements and the comparison to the mode matching approach is discussed in Zotter *et. al* [8]. As a first case study HOA of order 5 was implemented, resulting in a highly overdetermined system which can be solved by singular value decomposition (SVD). Future implementations will also include higher reproduction orders and more advanced decoders.

3. SYNCHRONIZATION

The audio rendering module including the decoder is implemented on four parallel processing units running Pure Data (Pd-0.42-5); each processing unit drives a subset of the 157 loudspeakers. Correct sound field reproduction requires time alignment of the loudspeaker driving signals with reference to the center of the room and synchronization of the computer output signals. The inter-machine communication is handled via IP using Open Sound Control (OSC) and User Datagram Protocol (UDP). In audio playback mode each renderer runs a sound file player, which is triggered and synchronized from the control computer (renderer 1). In real-time mode, the input audio signal is split and routed to each renderer to avoid additional system latency by audio I/O buffers.

The latency was measured using a Time Stretched Pulse (TSP) [9] signal, a variant of the swept-sinusoid. Deconvolution of the TSP signal from the recorded signals for each channel provides an impulse whose temporal position is equivalent to the delay. Using this characteristic, we examined the temporal synchronization. Signal acquisition was performed using an 8 channel audio interface (ProFire Lightbridge; M-Audio), A/D (ADA8000; Behringer) with recording software (Pro Tools 8 M-Powered; Desidesign). Using only 8 input channels, channel combinations were tested in batches in order to test all 157 channels. Fig 6. shows the system latency measurement process. The results show that the output signals from all D/A channels from the same PC were completely synchronous at the 1-sample level. However, the output signals between different PCs were not synchronous at the 1-sample level and the delay timing varies. The unsigned maximum latency was 51 samples (= 1.1 msec), requiring additional delays to compensate for this.

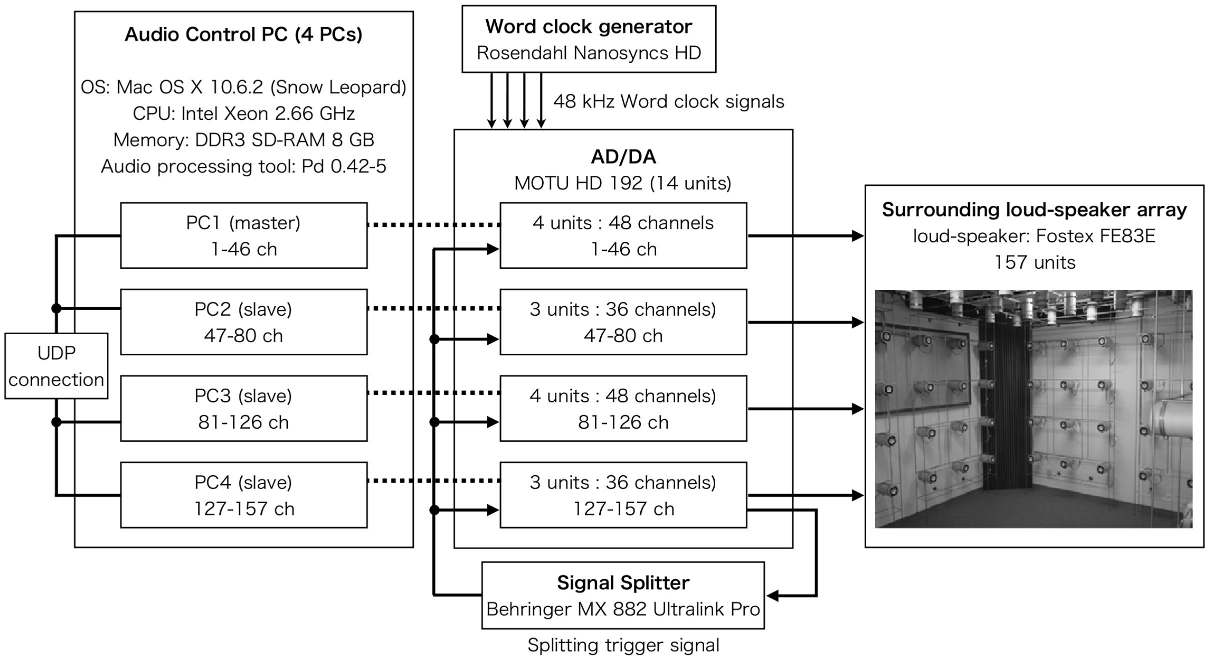

Figure 3: 157 loudspeaker array system

Figure 5: Appearance of the surroudnig loudspeaker array after installing the absorption materials

4. ACOUSTICS OF THE REPRODUCTION ENVIRONMENT

HOA reproduction requires free field conditions; because acoustic reflections and reverberation reduce the accuracy of the reproduced sound field. However, a free field environment is difficult to achieve, especially in a small room with many loudspeakers. The reverberation time in our reproduction room was initially about 0.2 s, which was far from free field. Therefore, additional sound absorption material was installed to allow more precise control of the sound field, especially for localization experiments and listening tests. Fig. 5 shows the room after installing the absorption materials. Room impulse responses were measured at the sweet spot before and after installing the sound absorption material (TSP, 1/2 inch microphone). The analysis of the results showed that the first reflections from each loudspeaker at the sweet spot ware reduced by about 10 to 20 dB. The reverberation time was reduced to a point where it could not be measured. By adapting a measurement method that is normally used for qualifying anechoic rooms, we found that the reproduction room was nearly free field within a 1 m radius from the sweet spot.

5. LOUDSPEAKER CHARACTERISTICS

In order to minimize reflections from the loudspeakers, small drivers are mounted into sealed cylindrical cardboard enclosures of minimal radius. The cylindrical enclosures contain sound absorptive material to minimise internal resonances. These loudspeakers are mounted in a regular grid parallel to the walls and the ceiling; this has the advantages of avoiding highly coherent reflections from the sweet spot and of optimally using the available space in the room. However, each loudspeaker has a directional response. In most HOA decoding systems, the loudspeakers are oriented to the sweet spot. In this system, the angle from the loudspeaker face varies, so correction of the response between each loudspeaker and the sweet spot is needed. Fig. 7 shows how the loudspeaker's magnitude response varies as a function of frequency and angle (based on measurements in an anechoic room). We are developing effective inverse filters for the correction from the impulse responses at the sweet spot. After calculating and applying the inverse filter to each loudspeaker, the final accuracy of the system's sound field reproduction can be assessed.

6. CONCLUSION

Although this paper presents a case study of a single HOA system, the problems and solutions discussed are relevant more generally to Ambisonics reproduction, especially in small rooms that are not initially anechoic.

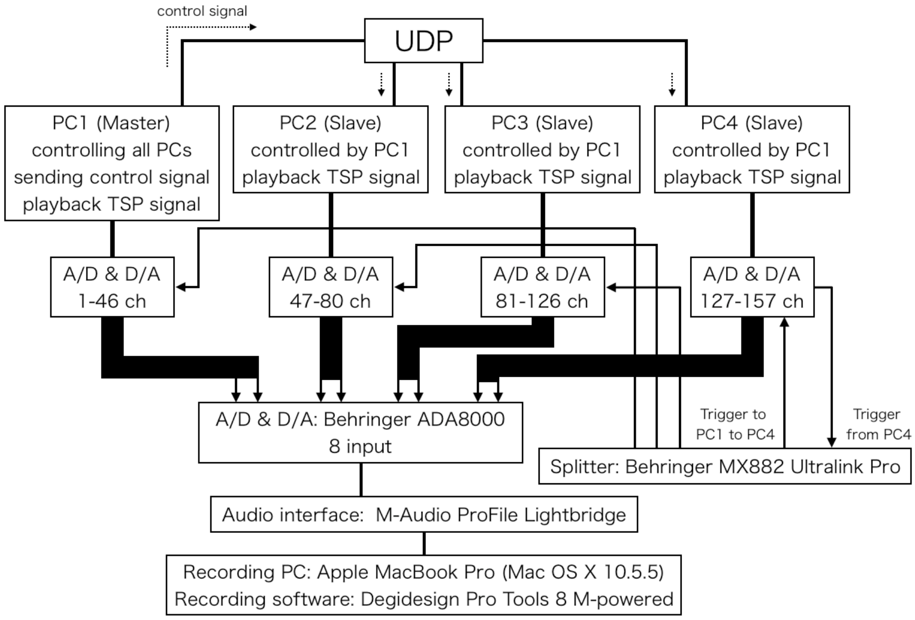

Figure 6: System latency measurement process

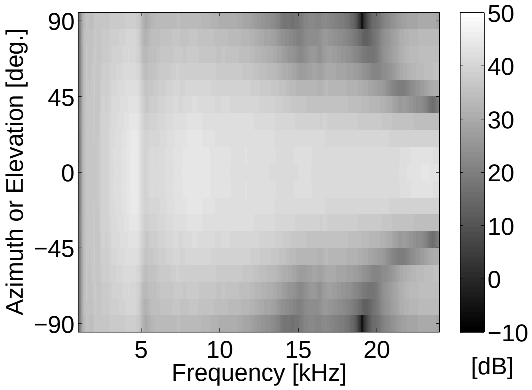

Figure 7: Amplitude-frequency characteristic of the loudspeaker (FE83E; Fostex) mounted in the cylindrical enclosure

Future work includes moving towards higher reproduction orders, applying advanced hemispherical Ambisonics decoder, sound field measurements and the perceptual evaluation of the overall system.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

[1] T. Okamoto, R. Nishimura, and Y. Iwaya, "Estimation of sound source positions using a surrounding microphone array," *Acoust. Sci & Tech.*, vol. 28, no. 3, pp. 181–189, May 2007.

[2] D.-G. Kang, Y. Iwaya, R. Miyauchi, and Y. Suzuki, "Comparison of sound localization performance between visual and real three-dimentional immersive sound field," *Acoust. Sci & Tech.*, vol. 30, no. 3, pp. 216–219, May 2009.

[3] T. Okamoto, B. F. Katz, M. Noisternig, Y. Iwaya, and Y. Suzuki, "Implementation of real-time room auralization using a surrounding 157 loudspeaker array," *Proc. IWPASH 2009 (eProceedings)*, no. 10.1142-9789814299312-0049, Nov. 2009.

[4] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[5] S. Ise, "A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems," *ACUSTICA - Acta Acustica*, vol. 85, no. 1, pp. 78–87, Jan. 1999.

[6] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.

[7] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. London, UK: Academic Press, 1999.

[8] F. Zotter, H. Pomberger, and M. Noisternig, "Ambisonic-decoding with and without mode-matching: case study using the hemisphere," *Proc. 2nd Int. Symp. on Ambisonics and Spherical Acoust.*, May. 2010.

[9] 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.*, vol. 97, no. 2, pp. 1119–1123, Feb. 1995.