PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Dynamic local sound field synthesis with multi-channel 1-bit signal reproduction system

Kakeru KUROKAWA(1), Izumi TSUNOKUNI(2), Yusuke IKEDA(3),Yasuhiro OIKAWA(4)

(1)Tokyo Denki University, Japan, 19fmi13@ms.dendai.ac.jp

(2)Tokyo Denki University, Japan, 19fmi23@ms.dendai.ac.jp

(3)Tokyo Denki University, Japan, yusuke.ikeda@mail.dendai.ac.jp

(4)Waseda University, Japan, yoikawa@waseda.jp

Abstract

To reproduce sound field accurately, local sound field synthesis(Local SFS) with virtual loudspeakers that are placed densely around a local listening area have been proposed. In local SFS, it is necessary to limit the size of the listening area. Thus, a listener's head can move out of a listening area due to the natural movement of the listener. In this paper, we propose the concept of a dynamic sound field synthesis system which can moves the sound field synthesis area so as to track the listener's head to allow the listener's movement. For the purpose of facilitating the construction of the massive loudspeaker system in an ordinary room, we utilize a playback system based on a high-speed 1-bit signal that can directly drive a loudspeaker. The proposed system can move the listening area by switching the corresponding local SFS signals obtained from the database including the source signals convolved with the driving functions to reproduce the multiple local area. In simulation experiments, reproduction errors of dynamic local and ordinary SFSs for a plane wave are compared to evaluate the accuracy of sound field by dynamic local SFS.

Keywords: Sound Field Reproduction, delta-sigma modulation, Digital Loudspeaker Array, CPU/FPGA SoC

1 INTRODUCTION

Recently, many techniques for physical sound field synthesis (SFS) have been studied such as wave field synthesis (WFS)[1, 2], higher-order ambisonics (HOA)[3, 4], boundary surface control (BoSC)[5], etc. In physical SFS, broadly arranging and controlling a large number of loudspeakers in a room is necessary. Thus, as the number of loudspeakers increases, the implementation cost rises. In our previous studies[6, 7, 8], the scale of reproduction system is small enough to construct a physical sound field reproduction system in a small room of an ordinary home, by utilizing a digital loudspeaker array which removes A/D converters and amplifiers.

To improve synthesis accuracy at frequencies higher than the spatial Nyquist frequency corresponding to the loudspeaker density, local SFS techniques have been studied[9]. To enable the listener to move freely in local SFS, the local reproduction area must track the movement/location listener's head.

In this paper, we introduce the concept of dynamic local SFS system using digital loudspeaker with multi-channel 1-bit signal. In addition, the reproduction errors of dynamic local SFS and ordinary SFS for a plane wave are compared to evaluate reproduction accuracy of dynamic local SFS by simulation experiments.

2 DYNAMIC LOCAL SOUND FIELD SYNTHESIS

Figure.1 shows the concept of dynamic local SFS. In general[10], the reproduced sound field $P(\mathbf{x},\omega)$ inside the region V enclosed by loudspeakers is represented as,

$$P(\mathbf{x},\omega) = \sum_{\mathbf{x}_0 \in \mathscr{X}_0} D(\mathbf{x_0},\omega)G(\mathbf{x_0},\mathbf{x},\omega) \quad (\mathbf{x} \in V), \tag{1}$$

where ω denotes the angular frequency, $D(\mathbf{x}_0,\omega)$ is driving function of a loudspeaker located at position $\mathbf{x}_0(\in \mathscr{X}_0)$, and $G(\cdot)$ a Green's function[10].In most physical SFS, the reproduction region is the entire region V enclosed by loudspeakers. However, in local SFS, the reproduction area is spatially limited to V_L to improve the reproduction accuracy at higher frequencies[9]. Many local SFS methods have been studied such as using focused sources, pressure-matching method, BoSC, etc.[5, 9, 11].

Now, consider the driving functions of loudspeakers to reproduce the sound field $S(\mathbf{x},\omega)$ in region V_L. In SFS using pressure-matching, BoSC method, or other pressure controlling methods, the driving function of loudspeaker x_0 $D(\cdot)$ is derived by

$$D(\mathbf{x}_0,\omega) = \sum_{\mathbf{x}\in V_L} S(\mathbf{x},\omega)G^{-1}(\mathbf{x}_0,\mathbf{x},\omega), \tag{2}$$

where $G^{-1}(\cdot)$ is the inverse matrix of Green's functions calculated by the least squares method,etc. Points $\mathbf{x}$ controlled by loudspeakers are referred to as control points. In the BoSC method, the controlled points $\mathbf{x}$ are restricted to the surface S_L based on Kirchhoff-Helmholtz equation.

In the LSFS with focused sources, the driving function is derived by

$$D(\mathbf{x}_0,\omega) = \sum_{\mathbf{x}_\mathrm{vss}\in \mathscr{X}_\mathrm{vss}} D_\mathrm{L}(\mathbf{x}_\mathrm{vss},\omega)D_\mathrm{vss}(\mathbf{x}_0,\mathbf{x}_\mathrm{vss},\omega), \tag{3}$$

where $\mathbf{x}_\mathrm{vss}(\in \mathscr{X}_\mathrm{vss})$ denotes the coordinate of the virtual secondary source, $D_\mathrm{L}(\cdot)$ is the driving function of the synthesizing virtual secondary source, and $D_\mathrm{vss}(\cdot)$ is the driving function of the virtual secondary source.

To change the reproduction area for tracking listener's head, the position of virtual secondary sources or control points must be changed. When the relative position between the desired sound source and reproduced area is changed by moving the sound source or reproduced area, both D_vss and $S(\mathbf{x},\omega)$ must be changed in Eqs.(2) and (3). In addition, both $D_\mathrm{L}(\mathbf{x}_\mathrm{vss},\omega)$ and $G^{-1}(\mathbf{x_0},\mathbf{x},\omega)$ must also be changed by moving the reproduced area. If the desired sound field is stable despite the spatial listening position such as a plane wave, then the change in D_vss and $S(\mathbf{x},\omega)$ is negligible at the amplitude of the reproduced area.

To implement dynamic local SFS, driving functions of loudspeakers are changed adjusting the listener's head position detected by a censor. To track the head in real time, the reproduced area must be moved to detected head position, before the head moves out of the area. In addition, the reproduced area should be moved with an enough margin, because of delay times of censor detection, convolution of driving function and the other processing times.

Unlike two-point sound pressure control techniques such as transaural system, local SFS can take the margin for real-time processing by expanding the reproduced area. However, reproduction accuracy and processing cost are changed by expansion of the reproduced area since the number of control points or virtual sources increases to keep the density of control points or virtual sources. Thus, the size of margin for the reproduced area should be decided in consideration of the trade-off between real-time factor and cost/accuracy of reproduction.

3 SYSTEM

3.1 Overview

Figure 2 shows a brief concept of the proposed dynamic local SFS system. In most techniques of local SFS, a large number of loudspeakers must be placed around a room. The proposed system consists of three types of units: a censor-unit to track a listener's head, a master-unit that manages all 1-bit signals, and hub-units for digitally driving loudspeakers.

3.1.1 Sensor Unit

The Sensor-Unit installed near the loudspeaker array detects the listener's head movement and then sends the coordinate of the listener's head to the Master-Unit. To achieve dynamical local SFS without any wearing devices,

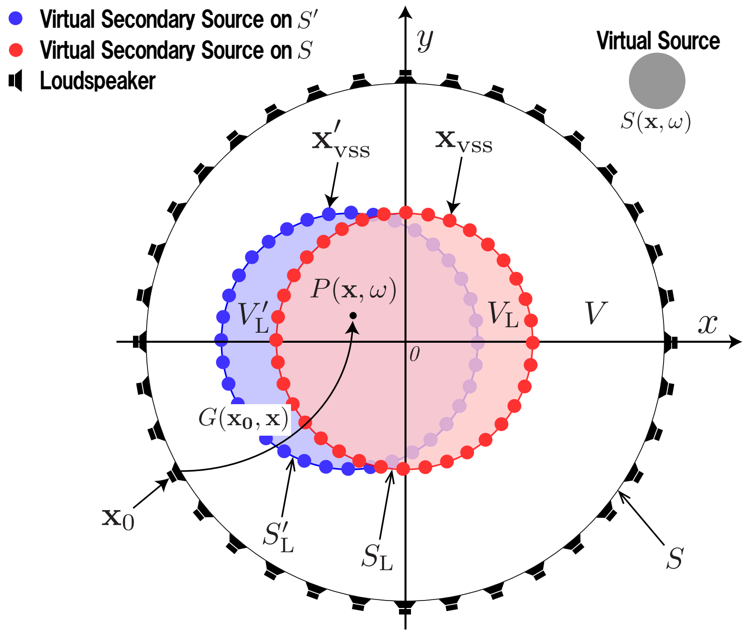

Figure 1. Concept of dynamic local sound field reproduction. $\mathbf{x}_0$ denotes the coordinate of the loudspeakers placed on the closed surface S; blue and red points are the coordinates of the virtual secondary source placed on the closed surface S_{L} and $S_{\mathrm{L}}{}'$; and $S(\cdot)$ is the sound field of the desired virtual source, S_{L} and S'_{L}.

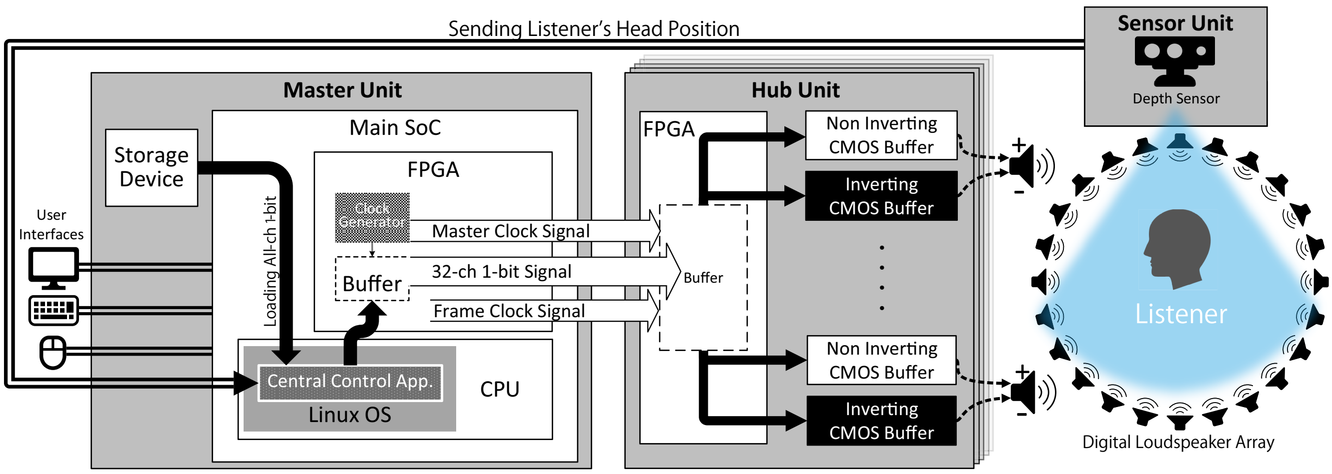

Figure 2. Brief concept of dynamic local SFS system.

non-contact depth sensors such as Microsoft Kinect or Intel RealSense is desirable to obtain the coordinate of the listener's head.

When the reproduction area is simply switched as shown in Fig.1, it is desirable that the reproduction area is updated until the center of head is distant about 20% of the diameter of the area from the center of area[12]. For example, a frame rate of Intel Real Sense D435i is 30 frame/sec. When the diameter of the reproduction area is 0.40 m and a distance of the listener's ears is 0.16 *m*, the available speed of listener's head tracked by reproduction area is about 2.4 *m*/sec without any delay time of processing to keep the distance between the head and the center of area within 0.08 m. Thus, when taking delay times of processing time into consideration, detection rate is high speed enough to track the listener walking inside the system.

3.1.2 Master Unit

A CPU/FPGA-integrated board is used for the main part of the Master-Unit. The FPGA is suitable for the signal processing which requires high-speed digital synchronization such as the generation of clock signals at frequencies of several megahertz, signal buffering, serializing, and dividing signals. The master unit receives the coordinate information of the listener's head from the Sensor-Unit involving the image processing with the external depth sensor. The CPU inside the main SoC requests the all-channels 1-bit signals from the storage device to synthesis sound field in the local reproduction area corresponding to the position of listener's head. The all-channels 1-bit signals are sent to the FPGA and then split into the serialized 32-ch 1-bit signals to send it to the each Hub-Unit. The serialized 32-ch 1-bit signals and two types of clocks: a master-clock and a frame-clock were sent to each Hub-Unit in differential form via three twisted-pair cables, such as a shielded CAT5e LAN cable which consisted of four twisted-pair cables.

3.1.3 Hub Units

Each Hub-Unit receives a master clock, a frame clock, and a serialized 32-ch 1-bit from the Master-Unit. The FPGA divides a serialized 32-ch 1-bit signal into each single-channel high-speed 1-bit signal based on the master-clock and the frame-clock. Then, the single channel 1-bit signal is sent to non-inverted and inverted CMOS buffers for increasing drive voltage of loudspeaker and remove a DC component. The loudspeaker is directly driven by the 1-bit signal since it can be restored to the original analog signal by passing the high speed 1-bit signal through a low pass filter. This type of fully-digital loudspeaker system is called as digital loudspeaker array[]. However, typical CMOS buffers are not designed to drive loudspeakers, but in the case of physical SFS in an ordinary small room, wavefront control is enough possible without a power amplifier.

4 SIMULATION

The loudspeaker's driving function depends on the positional relationship between the loudspeaker and the local reproduction area as stated Eq.2. Thus, the accuracy of the local reproduction area may differ depending on its position. By simulation, the accuracies in different local reproduction area with pressure matching method are compared. In simulation the Sound Field Synthesis Toolbox for MATLAB(SFS Toolbox ver. 2.4.2)[13] is used. Table 1 lists the simulation condition. To evaluate reproduction accuracies in the different local area with two reproduction methods; matching points and VSS with time reversal method, reproduction error is defined by

$$\mathrm{Error}(\mathbf{x}_i, \omega) = 10\log_{10}\frac{|P_\mathrm{d}(\mathbf{x}_i, \omega) - P_\mathrm{s}(\mathbf{x}_i, \omega)|^2}{|P_\mathrm{d}(\mathbf{x}_i, \omega)|^2}, \tag{4}$$

where P_d and P_s denote the desired complex sound pressure and the estimated complex sound pressure, respectively. Figure 3 shows the map of reproduction error in different reproduction areas by using PM method. As shown in the upper side of Fig.3(a)–(d) for matching point method, the reproduction error is almost same despite of the position of reproduced area. However, the reproduction error was increased only when the center of the reproduction area was placed in the direction of the sound source as shown in Fig.3(e) for matching point method. On the other hand, reproduction error becomes worse in local SFS with time-reversal VSS, when the

Table 1. Simulation Condition

Diameter of circular loudspeaker array [m]	4.0
Number of loudspeakers	84
Loudspeaker interval [m]	0.15
Number of matching point	3000
Radius of reproduced area [m]	0.2
SFS method (Condition 1)	Pressure Matching
SFS method (Condition 2)	VSS with time-reversal method
Number of VSS	100
Plane wave propagation direction	Y-axis negative

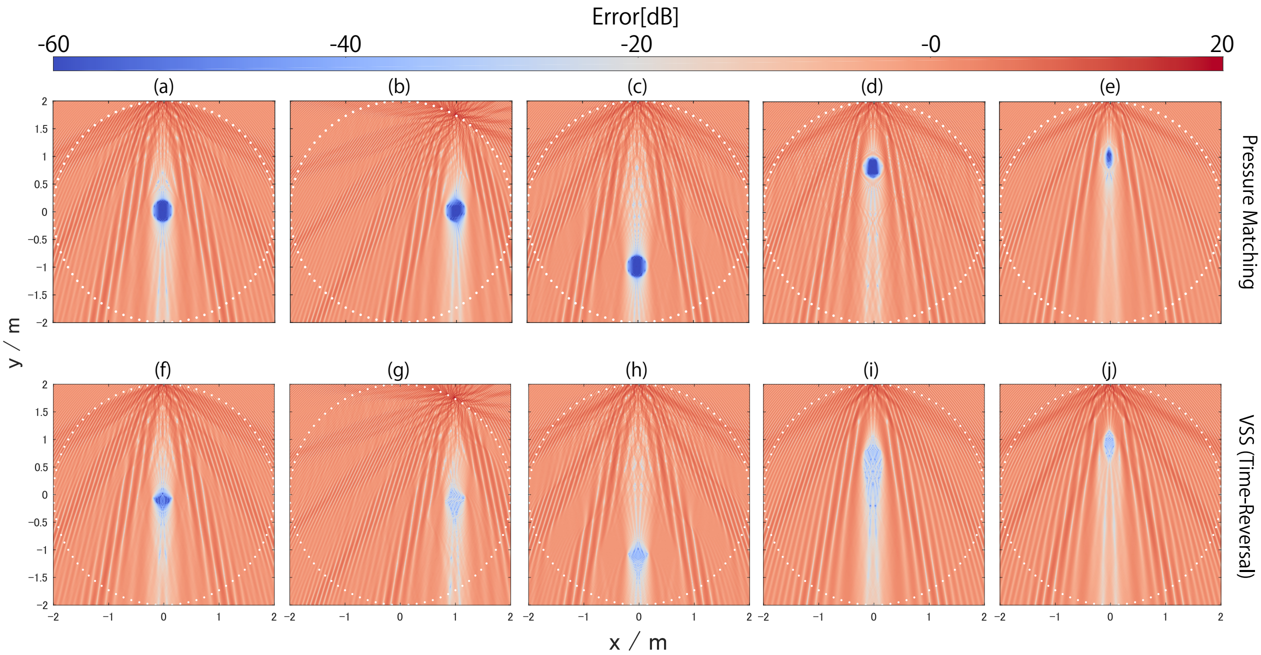

Figure 3. Reproduction errors at different reproduced area. Figs.(a)–(j) show the error of the local reproduction area at 8 kHz. From the right to the left, the center of reproduction area are [0.0 m, 0.0 m], [1.0 m, 0.0 m], [0.0 m, -1.0 m], [0.0 m, 0.8 m], [0.0 m, 1.0 m], respectively.

reproduction area is deviated from the center.

5 CONCLUSION

In this paper, we proposed the concept of dynamic local SFS system using a digital loudspeaker array and evaluate the accuracy of the dynamic local SFS by 2D simulation. The simulation results show that local SFS using PM method maintains higher accuracy at higher frequency by moving the position of the local reproduction area. In the future works, we will physically evaluate dynamic local SFS by using the proposed system.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 17KT0142.

REFERENCES

[1] A.J. Berkhout, D.D. Viries and P. Vogel, "Acoustic control by wave field synthesis", J. Acoust. Soc. Am. Vol.93, No.5, pp. 2764–2778, 1993.

[2] S. Spors, R. Rabenstein and J. Ahrens, "The Theory of Wave Field Synthesis Revisited", 124th AES Conv., 2008.

[3] P. Fellget, "Ambisonics. Part one: General system description", Studio Sound, Vol.17 pp. 20–22, 1975

[4] D.G. Malham and A. Myatt, "3D Sound Spatialization using Ambisonic Techniques", Computer Music Journal, Vol.9, No, 4, pp. 503–516, 1995.

[5] Shiro Ise, "A Principle of Sound Field Control Based on the Kirchhoff-Helmholtz Integral Equation and the Theory of Inverse Systems," Acta Acustica united with Acustica, Vol.85, No.1, pp. 78–87, 1999.

[6] Kurokawa et. al., "Three-Dimensional Large-Scale Loudspeaker Array System Using High-Speed 1-Bit Signal for Immersive Sound Field Reproduction," AES International Conf. on Spatial Reproduction, Jul. 2018.

[7] Yamanaka et. al., "Large-scale loudspeaker array system for sound field reproduction using high-speed 1bit signal processing," 5th ASA and ASJ Joint Meeting, Dec. 2016

[8] S. Pavan et. al., "Understanding Delta-Sigma Data Converters," Wiley-IEEE Press, 2016.

[9] Sascha Spors, J. Ahrens., "Local Sound Field Synthesis by Virtual Secondary Sources,"AES 40th International Conference, Oct. 2010.

[10] Jens Ahrens, "Analytic Methods of Sound Field Synthesis", Springer-Verlag Berlin Heidelberg, 2012.

[11] M.A.Poletti, "Improved methods for generating focused sources using circular arrays,"in AES 133rd convention, 2012.

[12] Kurokawa et. al., "Effect of Switching Reproduction Area in Dynamic Local" (submitted), IEEE 8th Global Conference on Consumer Electronics, 2019.

[13] H. Wierstorf, S. Spors, "Sound Field Synthesis Toolbox", in Proc. of 132nd Convention of the Audio Engineering Society, 2012.