

Powerful Sound Effects at Audience Seats by Wave Field Synthesis

Kimitaka Tsutsumi and Hideaki Takada

Abstract

Sound spatialization technology has become increasingly common within movie theaters and other venues for creating sound effects coming from all directions. NTT Service Evolution Laboratories has driven research into sound spatialization technology using a linear loudspeaker array to enhance the sense of reality in Kirari!, NTT's immersive telepresence technology. We have developed a system that enables the creation of sound images in front of loudspeakers that could not be created using conventional surround-sound systems. This article introduces the theory and implementation of wave field synthesis technology using a linear array of speakers for generating sounds that close in on venue seats.

Keywords: wave field synthesis, sound spatialization, spatial acoustics

1. Introduction

NTT Service Evolution Laboratories has been conducting research and development (R&D) of an immersive telepresence technology named “Kirari!” to replicate the reality of a sporting venue in real time at remote venues located anywhere in the world. Kirari! is a technology for showing a sports match as if it were happening in front of viewers by synchronously presenting virtual images and sound images of players consistent with their locations on a holographic display. In the case of soccer and judo, synchronous creation of sound images such as a kicked soccer ball and thrown judo players consistent with the locations of projected images provide higher reality to an extent that has been impossible using conventional stereo systems. For this reason, sound localization technology that creates virtual sound sources at any location within a venue is finding widespread use in movies and sports matches requiring a high sense of presence or powerful scenes. It is becoming essential for drawing out a highly realistic sensation using acoustics.

At NTT R&D Forum 2016, we used virtual speakers for sound localization by radiating ultrasonic waves in a controlled fashion at subjects on a screen. Since the audience would hear that sound reflected

off the screen, they would feel the sound as if it were coming directly from the subjects on the screen. However, control of the sound image was limited to left/right positions, which meant the results were hardly different from the sound produced by stereo reproduction. Another problem was that the use of reflected ultrasonic waves prevented the output of high sound pressure levels.

In response to these problems, we have investigated wave field synthesis technology capable of achieving sound-image control in the depth direction and consequently achieved sound reproduction with a high sense of presence in which sound can close in on venue seats. In addition, wave field synthesis can be implemented with standard speakers, so a large volume of sound can be produced relative to ultrasonic waves. This makes it possible to demonstrate a high-presence effect even in large venues. Wave field synthesis technology has been incorporated in Kirari! and successfully applied in the field at many events including NTT WEST Group Collection 2016, NTT R&D Forum 2017, Kabuki Virtual Theater, and Niconico Cho Kabuki 2017.

In this article, we present an overview of wave field synthesis technology and introduce the system used in the exhibition of this technology at Niconico Cho Kabuki 2017 as a field trial.



Fig. 1. Loudspeaker array.

2. Wave field synthesis technology

In this section, we explain the background of this technology and report on its application. We also describe problems that need to be addressed.

2.1 Theory of wave field synthesis

Wave field synthesis is sound-reproduction technology that reproduces a sound's wave front based on a physical model. It is achieved according to theory based on the Kirchhoff-Helmholtz integral equation [1]. This equation means that any sound field can be reproduced by controlling the sound pressure and sound-pressure gradient at the boundary surface enclosing that space. In essence, this means that if the walls, ceiling, and floor can be filled with infinitesimal speakers and if those speakers can be controlled, any sound field can be reproduced. However, if this were attempted using the Kirchhoff-Helmholtz integral equation in its original form and using commercially available speakers, it would require a large number of speakers with special characteristics that are difficult to achieve, and the total number of speakers needed would be massive.

We therefore decided on an implementation using an approximation based on Rayleigh's first integral and limiting the target sound field to a plane. In actuality, we form a speaker array with many closely spaced, linearly arranged speakers, set the target sound field, and synthesize it on the same plane as the speaker array (Fig. 1).

2.2 Kirari! and wave field synthesis technology

The wave field synthesis technology studied at the

NTT laboratories up to now should actually be called wave field reconstruction technology; this technology uses a speaker array to reproduce with high fidelity a sound field recorded with a microphone array [2]. Here, the sound field in the recording venue can be reproduced in its original form in the performance venue by processing and regenerating the recorded sound without having to detect the number and location of sound sources. However, because it is necessary to arrange the microphone array near the sound sources in the recording venue, the array may then appear in the video that is being synchronously recorded, making this setup difficult to use in an actual live broadcast. In addition, the difficulty here of separating individual sound sources detracts from the flexibility of limiting sounds to only specific ones and adding editing.

Considering, therefore, that direction and editing may be applied to content, we studied and adopted a new type of wave field synthesis technology that produces virtual sound sources from monaural sound sources. Editing and producing content after recording individual sound sources using a directional gun microphone as used in ordinary recordings and then adding sound spatialization effects to the monaural sound sources constitute a technique tailored to a realistic workflow.

2.3 Focused source method

Setting a target sound field so that the sound emitted from speakers creates a focus in front of the speaker array makes it possible to achieve a virtual sound source that pops up at this focus that is positioned nearer to the audience seats than the installed

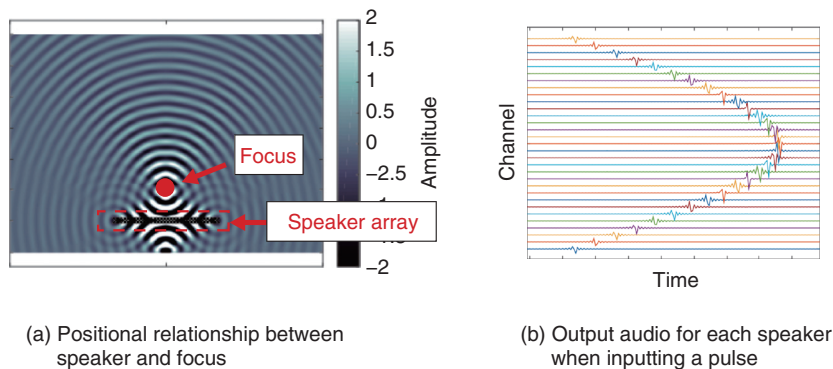


Fig. 2. Sound field creating a focused sound source and audio input to each speaker.

speakers. This sound-reproduction technique is called the focused source method [3].

When the focused source method is used, sound power can be concentrated at the point of focus by exploiting a digital filter determined analytically from the positions of the focus and speakers. Since the filters differ for each speaker, a monaural sound source is sufficient to obtain multichannel output even if using a large number of speakers. The role of the digital filter is to adjust the timing of sound reproduction and power at each speaker so that the sounds from every speaker arrive exactly at the same time at the focus position.

The positional relationship between the speaker array and the point of focus is shown in **Fig. 2(a)**. The audio output from each speaker in the case of a pulse input signal is shown in **Fig. 2(b)**. In this example, 32 speakers are lined up at 0.1-m intervals, and a focus is created 2 m from the center of the speaker array. Since the distance from either end of the speaker array to the focus is longer than the distance from the center of the speaker array to the focus, generating sound from these speakers simultaneously would cause the sound emitted from the center of the speaker array to pass the focus before the sounds emitted from both ends of the speaker array arrive at the focus. Consequently, to concentrate sound power at the focus, sound reproduction should begin at the speakers furthest away from the focus. On examining the output from each speaker shown in **Fig. 2(b)**, it can be seen that pulse generation begins at the speakers at each end of the array that lie furthest from the focus.

Next, we show the reproduced wave front when the sounds emitted from each speaker come together. If the sound source is placed at the focus position and a

single pulse is generated, the wave front will spread out in the shape of concentric circles (**Fig. 3**). The reconstruction of the desired output from the speaker array to generate a focus at the same position as above is shown in **Fig. 4**. It can be seen here that wave fronts similar to that shown in **Fig. 3** are formed where the sounds emitted from multiple speakers overlap. Using a digital filter determined by the focused source method as described above enables the generation of virtual sound sources in any location, including the area on the audience-seating side of the speaker array.

2.4 Problems in implementation

Unfortunately, using wave field synthesis does not enable sound effects that close in on the audience from the screen to be experienced from every seat in the venue. The theory of wave field synthesis assumes the use of an infinitely long speaker array, but this is not possible in practice since speaker arrays are actually formed by lining up a finite number of commercially available speakers. Furthermore, at an actual event, regulations such as fire laws or other constraints dictate where speakers can be installed, so it is not unusual to have to open up a gap for a passageway in a speaker array. This causes narrowing of the sweet-spot area in which the position at which an audience member perceives the sound image matches the position of a focused sound source set beforehand. As a result, the number of seats from which audience members can simultaneously experience the effects of pop-up sounds becomes smaller.

In response to this problem, we introduced a technique that creates a bend at a point in the linear speaker array to form a segmented array in which each segment acts as an independent array and that

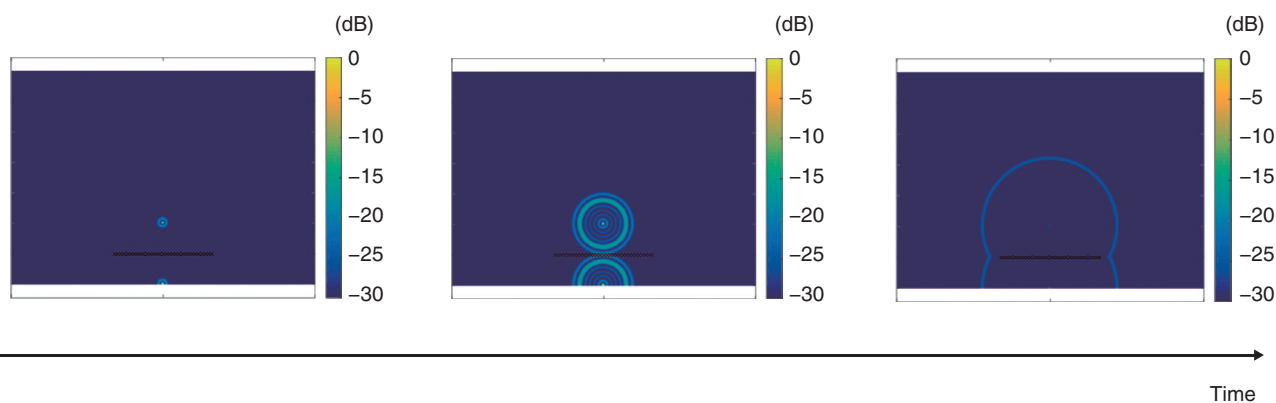


Fig. 3. Wave front of monopole sound source.

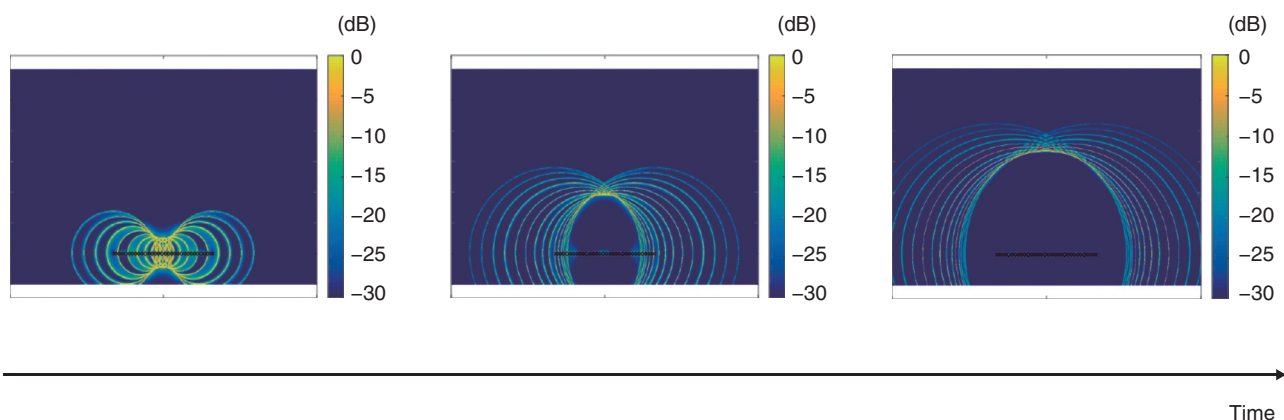


Fig. 4. Reproduced wave front using wave field synthesis.

generates a focused sound source for each segment, as shown in **Fig. 5** [4]. In this way, we succeeded in expanding the sweet-spot listening area. At present, research on this sweet-spot problem is progressing around the world, and we as well are working to further expand the sweet-spot listening area by pursuing theoretical advances in wave field synthesis.

3. Field trial: Niconico Cho Kabuki 2017

We introduced special effects at the Niconico Cho Kabuki 2017 event using 240 channels and speaker arrays with a total length of 36 m to perform wave field synthesis targeting 350 seats on the first level of the venue (**Fig. 6**). Specifically, we prepared sound effects achieved through wave field synthesis in a sound-reproduction personal computer (workstation) beforehand and had an operator activate these effects

in unison with the actor's performance as a system for reproducing sound effects.

The overall system configuration is shown in **Fig. 7**. This system sends out audio data from two audio interfaces connected in parallel at the workstation and reproduces audio from speakers with built-in amplifiers so that sound effects achieved by wave field synthesis resound in the venue. In this way, sound achieved through wave field synthesis was used to create effects for various scenes in the kabuki performance. For example, in the scene where a stone lantern explodes and a dragon flies away, we used the system to reproduce the sound of stone lantern fragments scattered by the explosion falling onto the audience at their seats, and in the final scene, we used it to reproduce the sound of a burning house collapsing near the audience. Generating the sound of falling debris or stones near the ears of audience members in

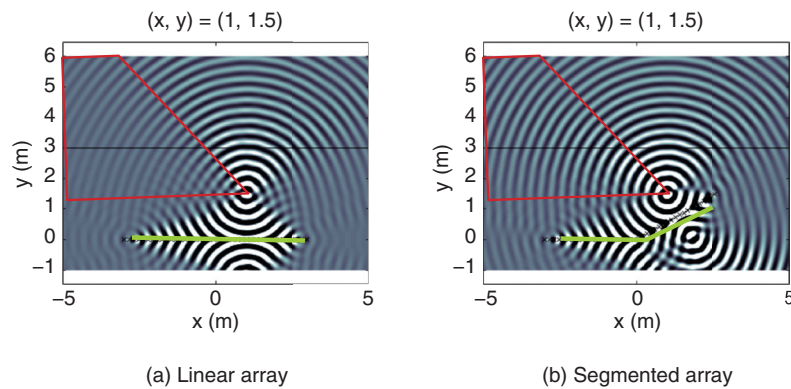


Fig. 5. Extended sweet spot. The area within the red border in (b) shows that the radiated sound from the focused source is stronger than the corresponding area in (a). This means that a larger area of the desired sound field is reproduced by a segmented array than by a linear array.

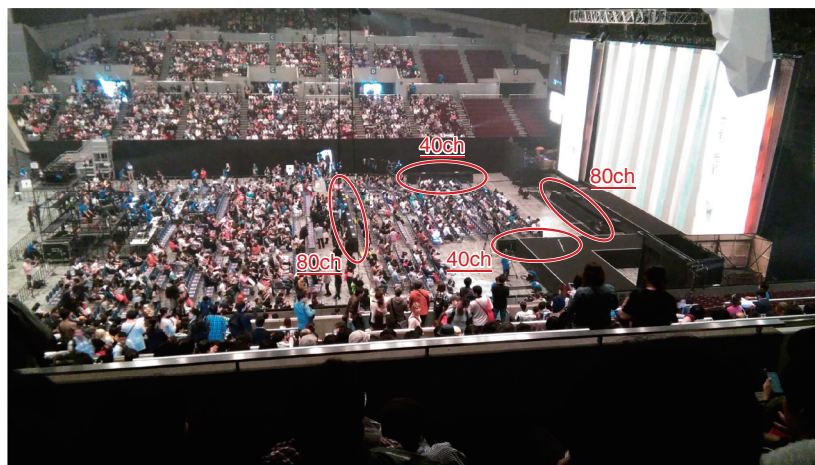


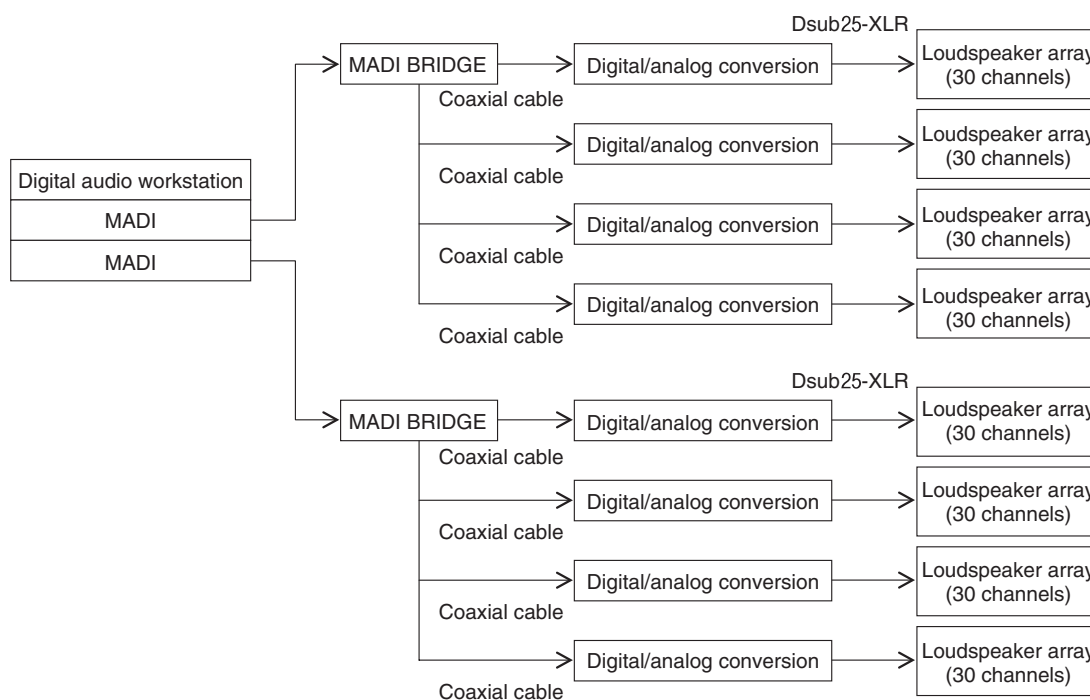
Fig. 6. Field trial at Niconico Cho Kabuki 2017.

this way achieved a high sense of presence as if debris was actually falling inside the venue.

Despite the fact that more than 5000 people attended each performance at this venue, Makuhari Messe Event Hall, constraints in installing equipment meant that only a portion of the audience seated on the first level were able to experience the special sound effects. Nevertheless, we succeeded in introducing special effects using wave field synthesis at a live event held in a large hall, and we received positive feedback on the effective and rich expression that we achieved in combination with the performance while dealing with the sweet-spot problem. This trial provided our team with an extremely valuable experience.

4. Future development

The use of a large number of speakers in systems applying wave field synthesis can create virtual sound images between loudspeakers and the audience not possible with surround-sound systems. However, there are still many problems to be addressed even with the use of a loudspeaker array, including the sweet-spot problem in the focused source method. It is therefore essential to continue researching wave field synthesis. Additionally, to provide home users with diverse content exploiting sound spatialization effects, we need to investigate sound spatialization technology using headphones or a small speaker array that can achieve the same effects as wave field



MADI: multichannel audio digital interface

Fig. 7. System configuration.

synthesis using a loudspeaker array. Here, research results in wave field synthesis are useful as benchmarks for such studies. By working on these problems in parallel, we will be able to research sound-reproduction technologies that achieve high reality for both theater and home use.

References

[1] K. Furuya and S. Koyama, "Wave Field Synthesis; Principles and Applications," Journal of the Institute of Image Information and Tele-

vision Engineers, Vol. 68, No. 8, pp. 621–624, 2014 (in Japanese).

- [2] S. Koyama, K. Furuya, H. Uematsu, Y. Hiwasaki, and Y. Haneda, "Real-time Sound Field Transmission System by Using Wave Field Reconstruction Filter and Its Evaluation," IEICE Trans. Fundamentals., Vol. E97-A, No. 9, pp. 1840–1848, 2014.
- [3] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis," Proc. of 127th AES Convention, paper 7914, New York, USA, Oct. 2009.
- [4] K. Tsutsumi and H. Takada, "A Method Using Auxiliary Arrays for Extending Listening Area of Focused Source in Wave Field Synthesis," Proc. of the 2017 Spring Meeting of the Acoustical Society of Japan, pp. 433–436, Kawasaki, Kanagawa, Japan, Mar. 2017 (in Japanese).



Kimitaka Tsutsumi

Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. and M.E. in informatics from Kyoto University in 2004 and 2006. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2006. He has participated in standardization efforts in International Telecommunication Union Telecommunication Standardization Sector (ITU-T). He joined NTT DOCOMO Research Laboratories in 2010, where he was involved in standardization activities of 3GPP (3rd Generation Partnership Project) on enhanced voice services. He has been with NTT Service Evolution Laboratories since 2015. His current research interests include speech coding and signal processing for spatial reproduction. He received the Technical Development Award from the Acoustical Society of Japan (ASJ) in 2016. He is a member of ASJ, the Audio Engineering Society, and the Institute of Electrical and Electronics Engineers (IEEE).



Hideaki Takada

Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. in information and communication technology from Tokai University, Shizuoka, in 1995, an M.E. in information systems from the University of Electro-Communications, Tokyo, in 1997, and a D.S. in global information and telecommunication studies from Waseda University, Tokyo, in 2007. He joined NTT Integrated Information & Energy Systems Laboratories in 1997. He has been researching three-dimensional (3D) visual perception and 3D display technology. He received the Best Paper Award from the 3D image conference in 2001 and 2005, the Young Researcher's Award from the Virtual Reality Society of Japan in 2002, the Achievement Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2003, a Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan in 2006, and Technical Committee Prize Paper Awards from IEEE in 2014 and 2016. He is a member of IEEE and the Institute of Image Information and Television Engineers.
