Acoust. Sci. & Tech. **41**, 1 (2020)

INVITED PAPER

Revisiting the theory of auditory displays based on the virtual sphere model

Jorge Treviño[1,2,*], Shuichi Sakamoto[1,2,†] and Yôiti Suzuki[1,2,‡]

[1]*Research Institute of Electrical Communication, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai, 980–8577 Japan*
[2]*Graduate School of Information Sciences, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai, 980–8577 Japan*

Abstract: Virtual Auditory Displays (VADs) are used to present realistic spatial sound. High-quality VADs must account for three factors: individuality (Head-Related Transfer Function), room acoustics (Room Transfer Function) and freedom of motion (active listening). The Auditory Display based on the VIrtual SpherE model (ADVISE) was proposed to simplify the problem by dividing it, through the Kirchhoff-Helmholtz integral theorem, into 1) a listener-free room acoustics simulation and 2) a free-field VAD using HRTFs. Users of ADVISE can move freely within the free-field region, thus accounting for active listening. This paper revisits the classic theory of ADVISE and identifies three oversights in the original proposal: 1) The ADVISE formulation suffers from non-unique boundary conditions at some frequencies. 2) The original proposal re-creates a set of boundary conditions using secondary sources that diverge on the boundary itself. 3) Considerations for sound propagation are absent in the original formulation. Two new formulations that retain the philosophy of ADVISE but are free from these problems are presented. The first one is based on the theory of Boundary Matching Filters, while the second is inspired by High-Order Ambisonics. The latter is found to be better suited for applications where freedom of motion is important since the presented sound field can be shifted by a translation matrix.

Keywords: High-order ambisonics, Binaural, Room acoustics, Auditory display

PACS number: 43.60.Dh, 43.60.Gk, 43.60.Sx [doi:10.1250/ast.41.276]

1. INTRODUCTION

The demand for realistic spatial sound has increased with the advent of virtual reality applications. Virtual auditory displays (VADs), and in particular headphone-based solutions, play an important role in conveying a natural and believable experience which allows the users of modern VR systems to become immersed in the presented environment. To achieve their goal, VADs often model the acoustic characteristics of the path between a given sound source and the listener's ears as a head-related transfer function (HRTF) [1]. The HRTF can realistically convey the illusion of sound arriving from a given direction; however, it varies between individuals and thus requires some degree of personalization.

A limitation of the HRTF is that, while it accounts for the acoustic effects due to the listener's body, it does not take into account the properties of the environment. Sound reflected from walls or obstructed by objects can drastically alter how we listen to an environment. The accurate re-creation of these effects is fundamental to achieve a believable experience leading to a high sense of presence. It is common to model environmental characteristics as a room transfer function (RTF) or, in combination with the HRTF, as a binaural room transfer function (BRTF). Unfortunately, the RTF and BRTF vary widely for different positions in a given environment.

Simple VADs may combine a single RTF with a set of HRTFs for multiple directions. This allows for the accurate presentation of direct sounds with the addition of coarse room reverberation effects. Modern virtual reality applications, however, give the users a high level of freedom to move around and explore the presented environment. This freedom of motion can considerably enhance the spatial perception of sound [2,3] if the VAD can render consistent auditory cues.

Summarizing, a realistic VAD must account for three fundamental factors: (1) the acoustic effects due to the listener's body, head and pinnae encoded in the HRTF, (2) the effects that reflecting surfaces and obstacles present in

*e-mail: jorge@ais.riec.tohoku.ac.jp
†e-mail: saka@ais.riec.tohoku.ac.jp
‡e-mail: yoh@riec.tohoku.ac.jp

the environment have on sound propagation, modeled by a RTF and (3) the ability of the listener to move around and explore the environment resulting in an accurate active listening experience. In theory, if the geometries and acoustic properties of the environment and listener's body are known, it is possible to calculate the sound pressure at the listener's ears due to an arbitrary sound source. Unfortunately this requires an impractically large computation making it unsuitable for real-time applications such as VR, where freedom of motion and interactivity are essential.

The problem can be divided into two simpler ones by making use of the Kirchhoff-Helmholtz integral theorem [4]. This observation led to the proposal of the Auditory Display based on the VIrtual SpherE model (ADVISE) [5,6]. The fundamental idea of ADVISE involves defining a spherical boundary that separates the listener from the environment (i.e. sound sources, walls and other objects that may affect sound propagation) as seen in Fig. 1. This way, the room acoustic characteristics can be calculated up to a fixed spherical boundary. Sound propagation inside the sphere obeys simple free-field conditions and, once the listener is introduced, are fully accounted for by a set of HRTFs covering all directions. The HRTFs can be swapped in real-time as the listener moves within the sphere, eliminating the need to update the room simulation part.

This paper revisits the theory of ADVISE as proposed in [5,6]. Section 2 summarizes the original formulation of ADVISE. While the idea of separating the problem through a boundary as described above is effective, some oversights are identified in Sect. 3. Finally, Sect. 4 applies the theory of acoustic holography, and in particular high-order Ambisonics (HOA), to overcome these issues.

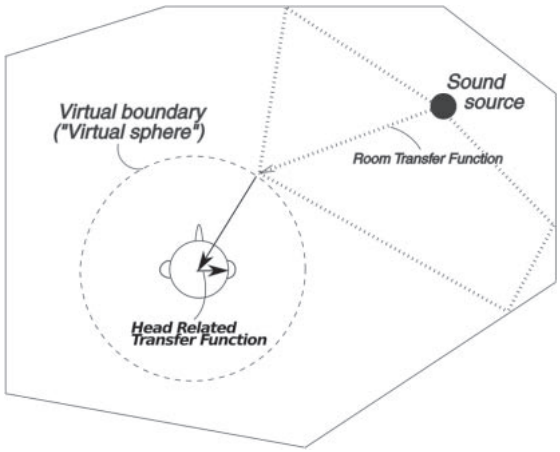

Fig. 1 Overview of the idea behind ADVISE. Sound propagation from the sound source, throughout the room and up to the listener's ears is modeled in two steps by introducing a spherical boundary. Room acoustic effects need to be calculated only outside of this boundary. On the other hand, free-field HRTFs are enough to render spatial sound anywhere inside the sphere. This allows the user to move around while retaining a stable spatial sound presentation without the need to continuously update the room acoustics model.

2. AUDITORY DISPLAY BASED ON THE VIRTUAL SPHERE MODEL

The Auditory Display based on the VIrtual SpherE model (ADVISE) seeks to present realistic spatial sound including room effects to a potentially moving listener. This requires real-time computation of the sound pressure at the listener's ears due to any number of sound sources, their reflections and scattering throughout the environment and over the listener's body, head and pinnae. To reduce computational costs, ADVISE introduces a spherical boundary which separates a free-field region, for the listener to move freely, from the environment. A room acoustics model can be solved outside the boundary through numerical methods such as finite-difference time-domain method (FDTD). This is computationally expensive, but only needs to be done once as long as the environment and the position of the boundary remain fixed. Free-field conditions inside the boundary make binaural rendering using a collection of HRTFs possible. These can be swapped easily in real-time as the listener moves within the sphere, providing accurate binaural signals for an active listening experience. The task of ADVISE is to match these two solutions at the boundary.

The original formulation of ADVISE proposes to treat the sphere as the boundary in the Kirchhoff-Helmholtz integral theorem. The theorem states that the sound pressure inside a bounded region is fully determined by the pressure and its derivative at the boundary [4]:

$$p(\boldsymbol{r},\omega) = \int_{\Gamma} \frac{\partial}{\partial \hat{n}}\, p(\boldsymbol{r}_\gamma,\omega) G(\boldsymbol{r}|\boldsymbol{r}_\gamma,\omega) - p(\boldsymbol{r}_\gamma,\omega)\, \frac{\partial}{\partial \hat{n}}\, G(\boldsymbol{r}|\boldsymbol{r}_\gamma,\omega) d\Gamma. \quad (1)$$

Here, the sound pressure p at a position $\boldsymbol{r}$ inside the boundary Γ is given for angular frequency ω as a surface integral. Position vectors $\boldsymbol{r}_\gamma$ lie on the boundary. The derivatives are taken with respect to the boundary surface's normal. The Green's functions G model sound propagation in the free-field and correspond to monopoles:

$$G(\boldsymbol{r}|\boldsymbol{r}_\gamma,\omega) = \frac{e^{-i\omega/c|\boldsymbol{r}-\boldsymbol{r}_\gamma|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}_\gamma|}. \quad (2)$$

We use the timelike sign convention so that solutions to the wave equation depend on $+ct - r$.

Numerical implementations require discretization of the boundary Γ in order to replace the integral of Eq. (1) by a sum. In addition to this, the original ADVISE proposal

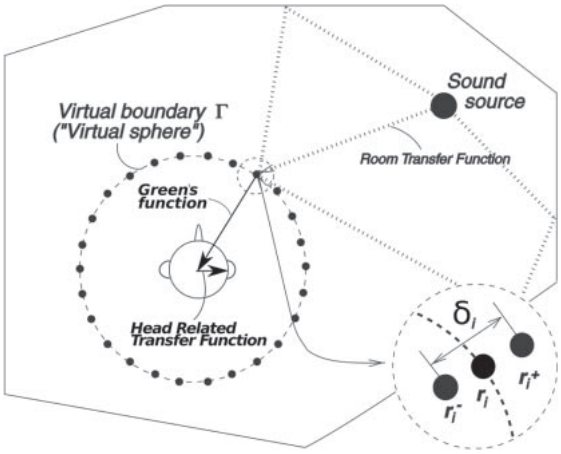

Fig. 2 The numerical implementation of ADVISE requires sampling the boundary surface and approximating the normal derivatives using a first-order difference.

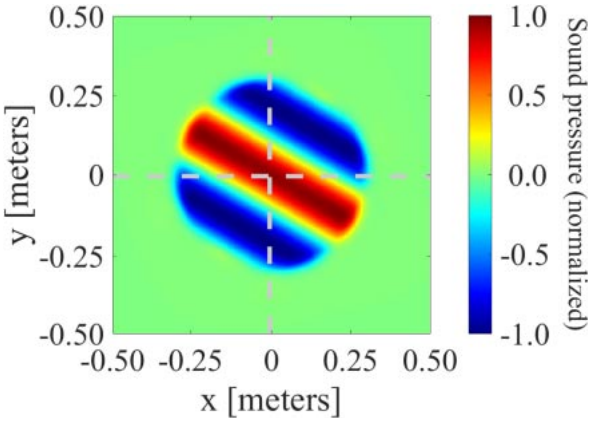

Fig. 3 Sound pressure field for a plane wave reproduced inside the ADVISE boundary. Computer simulations with properly tuned parameters appear to yield very high quality results at first glance.

relies on a first-order difference approximation to the normal derivatives. The approach is shown in Fig. 2 and summarized by the following equation:

$$p(\boldsymbol{r},\omega) \approx \sum_{i=1}^{N} \Bigg\{ G(\boldsymbol{r}|\boldsymbol{r}_i,\omega)[p(\boldsymbol{r}_i^+,\omega) - p(\boldsymbol{r}_i^-,\omega)] \frac{\Delta\Gamma_i}{\delta_i} - G(\boldsymbol{r}|\boldsymbol{r}_i^+,\omega) p(\boldsymbol{r}_i,\omega) \frac{\Delta\Gamma_i}{\delta_i} + G(\boldsymbol{r}|\boldsymbol{r}_i^-,\omega) p(\boldsymbol{r}_i,\omega) \frac{\Delta\Gamma_i}{\delta_i} \Bigg\}. \tag{3}$$

The boundary is sampled at N points Γ_i with quadratures $\Delta\Gamma_i$. The normal derivatives are approximated as the first-order difference between points $\boldsymbol{r}_i^+$ and $\boldsymbol{r}_i^-$, respectively outside and inside the boundary and separated by distance δ_i.

Binaural rendering from the formulation of Eq. (3) is achieved by multiplying the three Green's functions at $\boldsymbol{r}_i$, $\boldsymbol{r}_i^+$ and $\boldsymbol{r}_i^-$ by the corresponding HRTFs. The original ADVISE proposal does not detail any conditions for the sampling Γ_i; it simply takes the sound pressure calculated by the room acoustics model at the positions required by Eq. (3) and renders it binaurally. In this sense, it does not involve an inverse acoustic problem and is similar to some early attempts at reproducing sound fields by directly feeding microphone recordings into loudspeaker arrays [7].

3. PROBLEMS WITH THE ORIGINAL FORMULATION OF ADVISE

At first glance, ADVISE seems to deliver correct results as long as the boundary is sampled uniformly and the distance δ_i used to approximate the normal derivatives is properly chosen. An example of a plane wave being re-created inside the spherical boundary is shown in Fig. 3. There are, however, some problems with the formulation detailed in Sect. 2.

The first problem that Eq. (3) does not consider is related to the use of a spherical boundary to approximating the Kirchhoff-Helmholtz integral. This is known to lead to non-unique boundary conditions at specific wavenumbers $kr = m\pi$ with m being any integer [8]. This problem is present also in open microphone array recordings, such as those used in Boundary Surface Control (BoSC) [9]. A typical solution in sound field recording applications is to use a rigid baffle, as done in the SENZI binaural recording system [10] or to take multiple measurements with spheres of varying radii [11]. However, it is not immediately evident how could a rigid baffle or multiple concentric spheres may fit in the formulation of ADVISE.

Another problem is that, through Eq. (3), ADVISE attempts to re-create a finite sound pressure distribution over Γ using point sources located directly over it. Equation (2) shows that a collection of point sources on the boundary surface will instead result in an arbitrarily large sound pressure. This can be seen indirectly as the abrupt change in sound pressure that Fig. 3 exhibits at the boundary.

A final issue arises from how Eq. (3) in combination with the HRTF will simply render sound pressure measurements as virtual point sources. This approach disregards sound propagation since the effects that a point source at $\boldsymbol{r}_i$ have on other sampling positions $\boldsymbol{r}_j$ are ignored. The consequence of this is shown in Fig. 4, where the sound intensity vectors for a point source and the result of rendering by ADVISE differ substantially.

4. APPLICATION OF ACOUSTIC HOLOGRAPHY TO ADVISE

The theory of acoustic holography can be used to

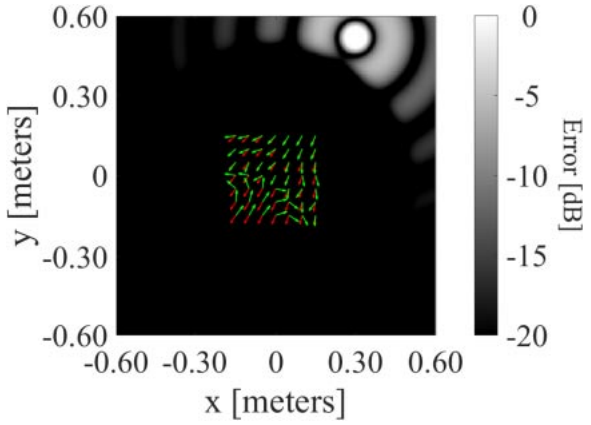

Fig. 4 Sound pressure reproduction error for a point source using ADVISE. Sound intensity vectors are shown in red for the ideal field and green for the reproduced field. While the error in sound pressure at a given time is considerably low, the intensity vectors show that sound does not propagate from the intended direction.

overcome the problems identified in Sect. 3. To this end, we go back to the basic task of matching two sound field models at a boundary surface.

Sound field simulations in this Section are for a monopole source radiating at a single frequency of 1 kHz. The source is located 1.5 meters away from the origin at an azimuth angle of 60 degrees. The virtual sphere boundary is set at a radius of 1 meter and sampled at the 252 directions used in the SENZI binaural recording system [10].

4.1. Boundary Matching Filters

An immediate solution is the use of the theory of boundary matching filters (BMFs) [12]. These are filters that transform sound pressure recordings on a boundary to single layer potentials on another. To avoid non-uniqueness of the boundary conditions, it is convenient to assume an acoustically rigid surface; this conditions are illustrated in Fig. 5. The choice calls for the use of *rigid to open* BMFs. The full procedure, described in [12], involves (1) the spherical harmonic decomposition of the measurements on the rigid sphere, (2) filtering each order's components using the respective BMF and (3) evaluation of the reproduction signals from the filtered spherical harmonic components.

The use of BMFs can eliminate all of the problems identified in Sect. 3. The rigid baffle prevents non-unique boundary conditions, the monopoles used to reproduce the recorded field do not lie on the recording surface and sound propagation is fully considered in the formulation of these filters.

Figure 6 shows the performance of a BMF-based version of ADVISE when reproducing a point source with

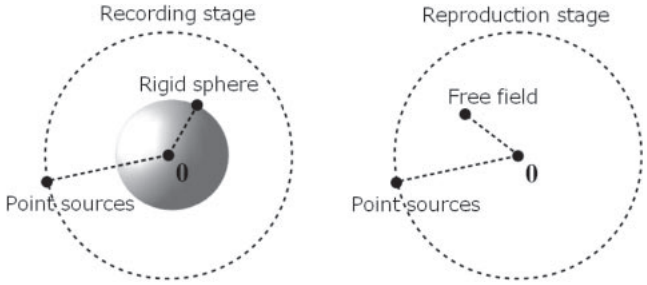

Fig. 5 Boundary matching filters can transform the sound pressure measurements over a boundary, such as a rigid sphere, into a single layer potential that can reproduce the recorded field from monopoles on a second boundary.

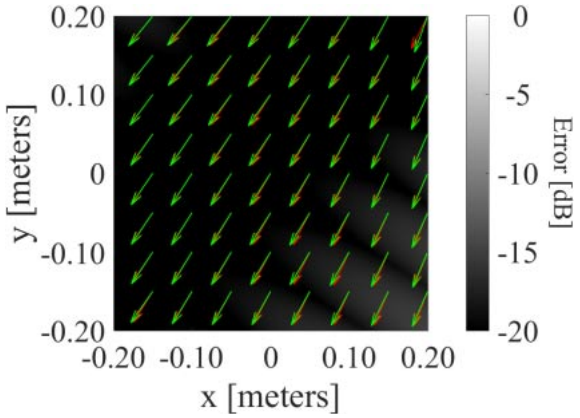

Fig. 6 Close up of the accurate reproduction region for a BMF-based ADVISE system presenting a point source. Red vectors show the ideal sound intensity, while green ones correspond to the reproduced field. While the region of accurate sound pressure reproduction is narrower than that of the original ADVISE proposal, ideal and reproduced sound intensity vectors are closely aligned.

the same parameters as the example of Fig. 4. The sound intensity vectors in the BMF-based revision of ADVISE show a maximum deviation from their ideal direction of under 4 degrees within 20 cm of the origin. Meanwhile, the error in their magnitude across the same region remained under -14 dB. Sound pressure is only accurate close to the center of the reproduction surface, as opposed to the original ADVISE formulation; however, within this region, the ideal and reproduced sound intensity vectors are correctly aligned.

4.2. High-order Ambisonics

One problem with using BMFs is the need for recordings on a rigid sphere. Since the sphere must be included in the room acoustics computation, its numerical complexity increases. The rigid sphere was introduced to remove non-unique boundary conditions. An alternative is to take sound field measurements at multiple concentric spheres [11].

However, the BMFs are not designed to match multiple boundaries into a single one.

An alternative is to use high-order Ambisonics (HOA) [13] to reproduce the recorded sound field. In this approach, multiple layer recordings are integrated into a single spherical harmonic representation which can then be decoded for a virtual loudspeaker array.

The spherical harmonic coefficients for an arbitrary sound field can be derived from free-field sound pressure measurements on a spherical boundary of radius a using the formula [13]:

$$B_{nm}(k) = \frac{1}{j_n(ka)} \int_{4\pi} p(a, \Omega, k) Y^*_{nm}(\Omega) d\Omega. \qquad (4)$$

Here, $k = \omega/c$ stands for the wavenumber. The integral covers all solid angles. Functions j_n and Y_{nm} are, respectively, the spherical Bessel and spherical harmonic functions of order n and degree m.

The division by the spherical Bessel functions in Eq. (4) may be undefined at certain frequencies and orders for which the functions are zero. This is similar to the issue identified in Sect. 3 when using an open recording boundary. However, this can be avoided by choosing the radius a appropriately so as to avoid the zeros. To avoid problems of numerical precision, it is possible to choose a for each wavenumber k and order n so as to maximize $j_n(ka)$.

The results of this approach are shown in Fig. 7. The sound intensity vectors in the HOA-based revision of ADVISE show a maximum deviation from their ideal direction of under 0.1 degrees within 40 cm of the origin. Meanwhile, the error in their magnitude across the same region remained under 30 dB. The region over which sound pressure is accurately reproduced is larger than that of the BMF-based approach, but still narrower than the original ADVISE implementation. Nevertheless, sound intensity vectors in the reproduced field are an almost perfect match to the ideal intensity vectors for the target sound source.

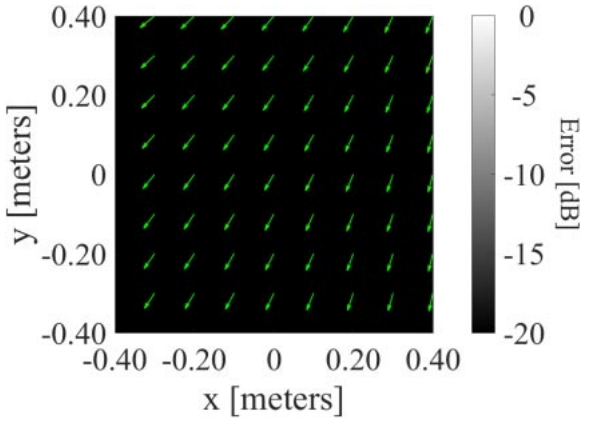

Fig. 7 Close up of the accurate reproduction region for a HOA-based ADVISE system presenting a point source. Ideal sound intensity vectors are hidden by the overlapping reproduction results, shown in green.

There is another advantage to the HOA-based approach. The coefficients obtained from Eq. (4) can be shifted using a translation matrix for spherical harmonic expansion coefficients [14]. When the listener moves in some direction, rather than updating the HRTFs used to render the binaural signals, it is possible to shift the sound field in the opposite direction. This way, the reproduction boundary can remain fixed, reducing the need to measure or compute a large number of HRTFs.

5. SUMMARY AND CONCLUSIONS

The present paper reviews the theory behind auditory displays based on the virtual sphere model. The original idea of joining a room acoustics model and a set of free-field HRTFs through a virtual boundary surface is a promising approach to rendering realistic spatial sound in real-time. However, the original implementation of ADVISE is not suitable for correct binaural presentation since it does not account for sound propagation between measurement and reproduction points.

Two alternative ADVISE implementations, inspired by acoustic holography were introduced. One relies on boundary matching filters and requires the inclusion of a rigid spherical boundary in the room acoustics simulation. The second proposal, based on high-order Ambisonics, does not have this requirement.

Both proposed approaches result in accurate sound pressure reproduction over a region that is narrower compared to the original ADVISE implementation; however, both present accurate sound intensity vectors pointing in the direction of intended sound propagation. The HOA approach seems to be the most promising as it results in larger listening regions. Further, it simplifies the task of updating the binaural signals to match listener movement.

ACKNOWLEDGEMENTS

Parts of this research were supported by the JSPS Grant-in-Aid for Scientific Research (KAKENHI) No. JP16H01736.

REFERENCES

[1] J. Blauert, *Spatial Hearing*, rev. ed. (MIT Press, Cambridge, Mass., 1996).
[2] Y. Iwaya, Y. Suzuki and D. Kimura, "Effects of head movement on front-back error in sound localization," *Acoust. Sci. & Tech.*, **24**, 322–324 (2003).
[3] Y. Suzuki, "Auditory displays and microphone arrays for active listening," *40th Audio Eng. Soc. Int. Conf.*, Keynote Address 2, p. 4 (2010).
[4] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, San Diego, 1999).

[5] S. Takane, Y. Suzuki and T. Sone, "A new method for global sound field reproduction based on Kirchhoff's integral equation," *Acta Acust.*, **85**, 250–257 (1999).
[6] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, "A new theory for high definition virtual acoustic display named ADVISE," *Acoust. Sci. & Tech.*, **24**, 276–283 (2003).
[7] M. Camras, "Approach to recreating a sound field," *J. Acoust. Soc. Am.*, **43**, 1425–1431 (1968).
[8] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," *J. Acoust. Soc. Am.*, **44**, 41–58 (1968).
[9] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *Acta Acust.*, **85**, 78–87 (1999).
[10] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya and Y. Suzuki, "Sound-space recording and binaural presentation system based on a 252-channel microphone array," *Acoust. Sci. & Tech.*, **36**, 516–526 (2015).
[11] F. M. Fazi, "Sound field reproduction," PhD thesis, University of Southampton (2010).
[12] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, "Boundary matching filters for spherical microphone and loudspeaker arrays," *IEEE/ACM Trans. Audio Speech Lang. Process.*, **26**, 461–474 (2018).
[13] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, **53**, 1004–1025 (2005).
[14] Y. Wang and K. Chen, "Translation of spherical harmonics expansion coefficients for a sound field using plane wave expansions," *J. Acoust. Soc. Am.*, **143**, 3474–3478 (2018).

Jorge Treviño graduated from the Monterrey Institute of Technology and Higher Education in 2005. He received the degree of M.Sc. in 2011 and a Ph.D. in information sciences in 2014, both from the Graduate School of Information Sciences of Tohoku University. He worked as an assistant professor in the Research Institute of Electrical Communication of Tohoku University from 2015 to 2019 and is currently a researcher in Yamaha Corporation. He is a member of member of AES, ASJ and IEICE and received the Awaya Kiyoshi young researcher award of the Acoustical Society of Japan in 2014. His research interests include sound field recording and reproduction, array signal processing, and spatial audio.

Shuichi Sakamoto received his B.S., M.Sc. and Ph.D. degrees from Tohoku University, in 1995, 1997, and 2004, respectively. He is currently a professor at the Research Institute of Electrical Communication, Tohoku University. He was a Visiting Researcher at McGill University, Montreal, Canada during 2007–2008. His research interests include human multi-sensory information processing including hearing, speech perception, and development of high-definition 3D audio recording systems. He is a member of ASJ, IEICE, VRSJ, and others.

Yôiti Suzuki graduated from Tohoku University in 1976 and received his Ph.D. degree in electrical and communication engineering in 1981. He is a professor emeritus of Tohoku University since 2019. His research interests include psychoacoustics, multi-modal perception, high-definition 3D auditory displays and digital signal processing of acoustic signals. He worked as the president of the Acoustical Society of Japan from 2005 to 2007. He received the Awaya Kiyoshi Award and three Sato Prizes from the society.