Acoust. Sci. & Tech. **38**, 2 (2017)

INVITED REVIEW

Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets

César D. Salvador*, Shuichi Sakamoto†, Jorge Treviño‡ and Yôiti Suzuki§

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

Abstract: Signal processing methods that accurately synthesize sound pressure at the ears are important in the development of spatial audio devices for personal use. This paper reviews the current methods and focuses on a promising class of these methods that rely on combining the spatial information available in microphone array recordings and datasets of head-related transfer functions (HRTFs). These two kinds of spatial information enable the consideration of dynamic and individual auditory localization cues during binaural synthesis. A general formulation for such a class of methods is presented in terms of a linear system of equations, whose associated matrix is composed of acoustic transfer functions that relate the positions of microphones and HRTFs. Based on this formulation, it is shown that most of the existing methods under consideration can be classified into two prominent approaches: 1) the HRTF modeling approach and 2) the microphone signal modeling approach. An important relation between these two approaches is evidenced in the general formulation: when one approach arises from the solution to an overdetermined system, the other corresponds to an underdetermined system, and vice versa. Illustrative examples of binaural synthesis from spherical arrays are provided by means of simulations. Underdetermined systems generally achieve better performance than overdetermined ones.

Keywords: 3D audio technology, Binaural technology, Head-related transfer functions, Microphone arrays, and spherical acoustics

PACS number: 43.60.Fg, 43.60.Uv, 43.66.Pn [doi:10.1250/ast.38.51]

1. INTRODUCTION

This paper concerns the recording of sound for its binaural reproduction. The idea behind this is as follows [1]: if sound pressure signals were recorded in the ears of a listener and reproduced exactly as they were, then the individual auditive experience is assumed to be re-created. This paper focuses on the generation of these signals, namely binaural signals that constitute the input to binaural reproduction devices based on headphones [2,3] or personal sound zone systems [4,5].

A classical binaural recording technique involves placing two microphones on the ears of a mock-up of the head and torso, namely a head and torso simulator (HATS). HATS recordings, however, have two major limitations. First, they do not reflect the movements of the listener's head, which are known to provide important dynamic cues for enhancing the perception of sound location [6–9]. Second, they do not consider the acoustic interactions with the individual traits of the listener's external anatomy before sound arrives to the eardrums. Such acoustic interactions can be described by the so-called head-related transfer functions (HRTFs), which are linear filters describing the transmission of sound from a position in space to the listener's eardrums [10,11]. Moreover, HRTFs are known to contain individual auditory cues that are important to present sounds accurately located outside of the head [2,8,12,13].

Updating binaural signals to match the listener's motion is possible by using a HATS that simultaneously moves as the listener does. This system is known as a TeleHead [14,15]. TeleHeads, however, can match the simultaneous motion of only one listener at a time, and motion can only be tracked at the same time of recording. Besides, the motion-tracking system of TeleHeads uses a complicated mechanical system. The tracking of multiple moving listeners is possible electronically when recording is performed with a microphone array. The use of micro-

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

phone arrays to capture spatial sound enables the tracking of movements by means of digital computations in real time or after recording. The first proposal of such a system was the motion-tracked binaural (MTB) system [16].

Including individual auditory cues into microphone array recordings would require the modeling of individual auditory spaces. This can be achieved by using individual HRTFs obtained for a set of positions in space. Recent advances in technologies for measurement [17] and computation [18–25] are now allowing us to obtain individual HRTFs for dense sets of positions. Moreover, individual HRTF datasets are available in public databases reported in [26–29].

During the last decade, researchers have shown an increasing interest in a modern class of methods for obtaining the binaural signals by combining microphone array recordings with individual HRTF datasets [30–56]. The electronic processing of the microphone array outputs in correspondence with the listener's head movements, and the combination of the results with an individual HRTF dataset, aim to generate the dynamic auditory cues individually, without the need to physically move a HATS or the microphone array itself.

The purpose of this paper is to review studies on such modern binaural recording systems and present a general formulation and classification for them. The remainder of the paper is organized as follows. Section 2 reviews past studies to elucidate the difference between the widely known three-dimensional auditory displays and the framework for binaural recording and reproduction that demarcates this paper. Section 3 presents a general formulation and classification for modern binaural synthesis methods that combine microphone array recordings and HRTF datasets. Section 4 describes methods to solve the underlying linear systems by decoupling the positions of microphones and HRTFs. Section 5 describes a physical model for use with spherical arrays. Section 6 presents simulations of binaural synthesis from spherical arrays. Section 7 summarizes this paper.

2. THREE-DIMENSIONAL AUDITORY DISPLAYS AND BINAURAL SYSTEMS

Auditory displays are devices to present sound in general. When the spatial features of sound are included, these devices are referred to as three-dimensional (3D) auditory displays [57,58]. Spatial sound can be presented in the form of binaural signals by means of HRTFs [2]. This type of systems, i.e., binaural reproduction systems, are a class of 3D auditory displays. When binaural reproduction systems are used as 3D auditory displays to render virtual sound, a typical input is a monophonic signal obtained through recording or artificial generation. In this case, the binaural signals for a target position are calculated by simply convolving the monophonic signal with the time-domain representation of the HRTFs, the so-called head-related impulse responses, obtained beforehand for the target position.

To calculate binaural signals for virtual sounds at arbitrary target positions, it is usual to use HRTFs obtained beforehand for a set of initial positions. The initial HRTF dataset is typically obtained for an array of positions at a single distance from the head center [27–29]. From this initial dataset, the HRTFs for arbitrary target positions that are not contained in the initial dataset are synthesized by interpolation along angle [59–63] or extrapolation along radial distance [64–67].

The binaural systems considered in this paper, on the other hand, aim to reconstruct the binaural signals due to real acoustic environments. Binaural signals are seen as a combination of the sound field due to an acoustic environment with the model of an individual auditory space. Sound fields are captured by means of microphone arrays, whereas individual auditory spaces are modeled by HRTF datasets. In this context, an HRTF dataset is regarded as a directivity pattern due the interactions of sound with the body, head, and pinna.

A general formulation and classification for a modern class of binaural recording systems based on combining microphone array recordings and HRTF datasets is presented in the next section.

3. BINAURAL SYNTHESIS FROM MICROPHONE ARRAY RECORDINGS AND HRTF DATASETS

During the last decade, several binaural synthesis methods that combine the spatial information contained in microphone array recordings with the spatial information contained in individual HRTF datasets have been studied [30–56]. The use of a microphone array offers the possibility of including the dynamic auditory cues used in sound localization because the movements of multiple listeners can be tracked through digital computations in real-time or non-real-time conditions. The HRTF datasets allow the inclusion of the auditory localization cues that correspond to the individual traits of the listeners, provided that individual HRTF datasets can be obtained directly for each listener or by means of anthropometric personalization [68,69] or individualization methods [70,71]. Moreover, HRTFs for dense sets of positions can be calculated based on interpolation methods [59–67]. In general, this class of binaural synthesis methods aim at rendering a sound pressure field sampled at the positions where a set of HRTFs is given. In other words, this class of methods aim to synthesize the binaural signals due to an array of virtual loudspeakers placed at the positions used to obtain the HRTFs.

It is argued in this section that most of the existing methods for combining microphone array recordings with HRTF datasets can be classified into two prominent approaches: 1) the HRTF modeling approach and 2) the microphone signal modeling approach. Each approach refers to the manner in which the underlying inverse problem is formulated.

In the HRTF modeling approach, an HRTF dataset constitutes a specified spatial pattern to be approximated by a set of weighting filters applied to the microphone recordings. The weighting filters are calculated by solving a linear system of equations that approximate the HRTF dataset. The entries of the matrix associated to the linear system are acoustic transfer functions from the positions of microphones to the positions used to obtain the HRTF dataset. Examples of binaural systems based on this approach are the SENZI system [35–39], the virtual artificial head (VAH) [41–44], binaural beamforming systems [34,45,47,49,52,53], and the implementations reported in [40,56].

In the microphone signal modeling approach, on the other hand, the microphone array recordings are used to calculate the driving signals at the positions used to obtain the HRTF dataset. The driving signals are subsequently rendered with the corresponding HRTFs in the dataset by relying on the principle of acoustic wave superposition [72]. The driving signals are calculated by solving a linear system of equations that model the microphone array recordings. The entries of the associated matrix are acoustic transfer functions from the positions involved in the HRTF dataset to the positions of microphones. Examples of binaural systems based on this approach are the binaural system obtained when combining the BPLIC [30] and ADVISE systems [31,32], and the binaural ambisonic systems treated in [33,46,48,50,51,54,55].

To support the assertions above, a general formulation for the two approaches is proposed below in this section. A formulation in continuous space is described in Sect. 3.1, whereas a formulation in discrete space is detailed in Sect. 3.2. A remarkable relation between the HRTF modeling approach and the microphone signal modeling approach will be evidenced in the discrete-space formulation: when one approach arises from the solution to an overdetermined system of equations (more equations than unknowns), the other corresponds to an underdetermined system of equations (less equations than unknowns), and vice versa. This relation is overviewed in Table 1.

In all of what follows, acoustic pressure signals and transfer functions are assumed to have finite energy. They are represented in the frequency domain by complex-valued functions that are square integrable. For simplicity, the variation of functions with frequency is not explicitly indicated.

Table 1 Two approaches for binaural synthesis.

Condition	Approach	
	HRTF modeling	Microphone signal modeling
Number of microphones less than number of HRTFs ($M < L$)	Overdetermined system	Underdetermined system
Number of microphones greater than number of HRTFs ($M > L$)	Underdetermined system	Overdetermined system

3.1. Continuous-space Formulation

The aim is to show that combining recordings p and HRTF datasets $h = \{h^{\text{left}}, h^{\text{right}}\}$ to synthesize the binaural signals $\varphi = \{\varphi^{\text{left}}, \varphi^{\text{right}}\}$ can be formulated in terms of an inverse problem. It is assumed that p is known at every position $\vec{a}$ in a region of space $\mathcal{A}$, while h was obtained for every position $\vec{b}$ in a region of space $\mathcal{B}$. It is further assumed that $\mathcal{A}$ and $\mathcal{B}$ are continuous, arbitrarily shaped, and non-intersecting regions.

More specifically, the aim is to show that the two approaches for binaural synthesis can be formulated in terms of the following expression:

$$\varphi = \int_{\vec{b}\in\mathcal{B}} \int_{\vec{a}\in\mathcal{A}} \overline{h(\vec{b})} C^{\dagger}(\vec{a},\vec{b}) p(\vec{a}) d\vec{a} d\vec{b}. \tag{1}$$

The overbar symbol is used to denote the complex conjugate hereafter. By C, we denote the transmission of sound between $\vec{a}$ and $\vec{b}$. This acoustic transfer function also takes into consideration the geometrical and physical properties of $\mathcal{A}$ and $\mathcal{B}$. It is assumed that C can be obtained beforehand. The conditions for the existence of its inverse, denoted by $C^{\dagger}$, are investigated below for each binaural synthesis approach. As for the statement of the corresponding inverse problems, integration over $\mathcal{A}$ of an acoustic transfer function C_{HRTF} from $\vec{a}$ to $\vec{b}$ will be used in the HRTF modeling approach, whereas integration over $\mathcal{B}$ of an acoustic transfer function C_{mic} from $\vec{b}$ to $\vec{a}$ will be used in the microphone signal modeling approach.

3.1.1. HRTF modeling approach

The binaural signals φ are synthesized by applying weighting filters $w = \{w^{\text{left}}, w^{\text{right}}\}$ to the microphone array recordings p in such a way that

$$\varphi = \int_{\vec{a}\in\mathcal{A}} \overline{w(\vec{a})} p(\vec{a}) d\vec{a}, \tag{2}$$

where the weighting filters are defined by

$$w(\vec{a}) = \int_{\vec{b}\in\mathcal{B}} C^{\dagger}_{\text{HRTF}}(\vec{a},\vec{b}) h(\vec{b}) d\vec{b}. \tag{3}$$

It can be verified that equating (3) and (2) results in an expression similar to (1) if $\mathcal{C}^{\dagger} = \mathcal{C}^{\dagger}_{\mathrm{HRTF}}$.

Equation (3) is the result of solving an inverse problem that is formulated by modeling the HRTF dataset h as an integral of transfer functions $\mathcal{C}_{\mathrm{HRTF}}$ as follows:

$$\int_{\vec{a}\in\mathcal{A}} C_{\mathrm{HRTF}}(\vec{a},\vec{b})w(\vec{a})d\vec{a} = h(\vec{b}). \tag{4}$$

A procedure to solve (4) relies on postulating a trial transfer function $C^{\dagger}_{\mathrm{HRTF}}$ from $\vec{b}$ to arbitrary $\vec{a}'$ that is used to multiply both sides of (4) and integrate over $\mathcal{B}$. A further change of the order of integration in the double integral results in

$$\begin{aligned}\int_{\vec{a}\in\mathcal{A}} \left\{ \int_{\vec{b}\in\mathcal{B}} C^{\dagger}_{\mathrm{HRTF}}(\vec{a}',\vec{b})C_{\mathrm{HRTF}}(\vec{a},\vec{b})d\vec{b} \right\} w(\vec{a})d\vec{a} \\ = \int_{\vec{b}\in\mathcal{B}} C^{\dagger}_{\mathrm{HRTF}}(\vec{a}',\vec{b})h(\vec{b})d\vec{b}.\end{aligned} \tag{5}$$

If the condition

$$\int_{\vec{b}\in\mathcal{B}} C^{\dagger}_{\mathrm{HRTF}}(\vec{a}',\vec{b})C_{\mathrm{HRTF}}(\vec{a},\vec{b})d\vec{b} = \begin{cases} 1 & \text{if } \vec{a} = \vec{a}', \\ 0 & \text{else}, \end{cases} \tag{6}$$

holds, then substituting (6) in (5) yields (3). The trial transfer function $C^{\dagger}_{\mathrm{HRTF}}$ thus defines an inverse for C_{HRTF} under the condition for invertibility (6).

3.1.2. Microphone signal modeling approach

The binaural signals φ are synthesized in this case by using the HRTF dataset h to render a set of driving signals u according to the following expression:

$$\varphi = \int_{\vec{b}\in\mathcal{B}} \overline{h(\vec{b})}u(\vec{b})d\vec{b}, \tag{7}$$

where the driving signals are defined by

$$u(\vec{b}) = \int_{\vec{a}\in\mathcal{A}} C^{\dagger}_{\mathrm{mic}}(\vec{a},\vec{b})p(\vec{a})d\vec{a}. \tag{8}$$

It can be verified that equating (8) and (7) results in an expression similar to (1) if $C^{\dagger} = C^{\dagger}_{\mathrm{mic}}$.

Equation (8) arises from solving an inverse problem that is formulated by modeling the microphone recordings p as an integral of transfer functions C_{mic} in such a way that

$$\int_{\vec{b}\in\mathcal{B}} C_{\mathrm{mic}}(\vec{a},\vec{b})u(\vec{b})d\vec{b} = p(\vec{a}). \tag{9}$$

Equation (9) is solved by assuming a trial transfer function $C^{\dagger}_{\mathrm{mic}}$ from $\vec{a}$ to arbitrary $\vec{b}'$ that is used to multiply both sides of (9) and integrate the product over $\mathcal{A}$. Changing the order of integration in the double integral results in

$$\begin{aligned}\int_{\vec{b}\in\mathcal{B}} \left\{ \int_{\vec{a}\in\mathcal{A}} C^{\dagger}_{\mathrm{mic}}(\vec{a},\vec{b}')C_{\mathrm{mic}}(\vec{a},\vec{b})d\vec{a} \right\} u(\vec{b})d\vec{b} \\ = \int_{\vec{a}\in\mathcal{A}} C^{\dagger}_{\mathrm{mic}}(\vec{a},\vec{b}')p(\vec{a})d\vec{a}.\end{aligned} \tag{10}$$

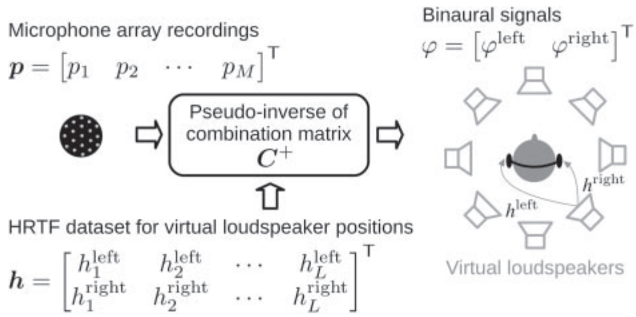


Fig. 1 Overview of binaural systems.

It is verified from (10) that if the condition

$$\int_{\vec{a}\in\mathcal{A}} C^{\dagger}_{\mathrm{mic}}(\vec{a},\vec{b}')C_{\mathrm{mic}}(\vec{a},\vec{b})d\vec{a} = \begin{cases} 1 & \text{if } \vec{b} = \vec{b}', \\ 0 & \text{else}, \end{cases} \tag{11}$$

holds, then replacing (11) in (10) results in (8). The trial transfer function $C^{\dagger}_{\mathrm{mic}}$ thus defines an inverse for C_{mic} under the condition for invertibility (11).

3.2. Discrete-space Formulation

Practical implementations of the models presented in Sect. 3.1 require a discrete-space version in terms of matrix computations. To this aim, a finite number M of microphones and a finite number L of HRTFs are considered in this section. Signals are assumed in the frequency domain. An overview of the discrete-space structures under consideration is illustrated in Fig. 1.

The aim is to show that binaural synthesis can be formulated in terms of the following discrete-space version of (1)

$$\varphi = \overline{\boldsymbol{h}^{\top}}\boldsymbol{C}^{+}\boldsymbol{p}. \tag{12}$$

Here, the synthesized binaural signals for the left and right ears are organized in

$$\varphi = \begin{bmatrix} \varphi^{\mathrm{left}} & \varphi^{\mathrm{right}} \end{bmatrix}^{\top}. \tag{13}$$

The symbol $^{\top}$ indicates transpose.

The recordings of an array composed of M microphones are organized in the vector

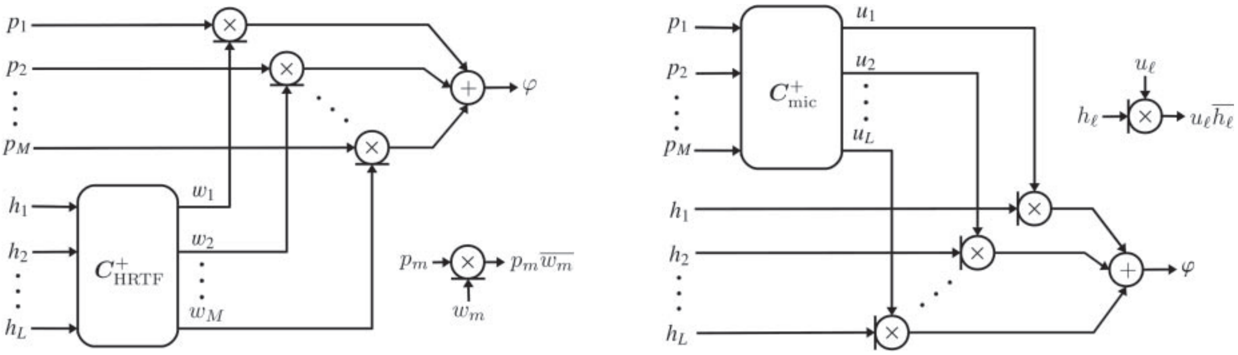
$$\boldsymbol{p} = \begin{bmatrix} p_1 & p_2 & \cdots & p_M \end{bmatrix}^{\top}. \tag{14}$$

Each entry p_m of $\boldsymbol{p}$, where $m = 1, 2, \ldots, M$, represents a sample of sound pressure recorded at an arbitrary position $\vec{a}_m$. The finite set $\{\vec{a}_m\}_{m=1,2,\ldots,M}$ is a sampling of the region of space $\mathcal{A}$ where $\boldsymbol{p}$ is obtained.

The HRTFs of the dataset are organized in the matrix

$$\boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}^{\mathrm{left}} \\ \boldsymbol{h}^{\mathrm{right}} \end{bmatrix}^{\top} = \begin{bmatrix} h_1^{\mathrm{left}} & h_2^{\mathrm{left}} & \cdots & h_L^{\mathrm{left}} \\ h_1^{\mathrm{right}} & h_2^{\mathrm{right}} & \cdots & h_L^{\mathrm{right}} \end{bmatrix}^{\top}. \tag{15}$$

Each entry h_{ℓ}^{left} or $h_{\ell}^{\mathrm{right}}$ of $\boldsymbol{h}$, where $\ell = 1, 2, \ldots, L$, represents a sample of the free-field HRTF for the left or



(a) HRTF modeling. (b) Microphone signal modeling.

Fig. 2 Two approaches for binaural synthesis that combine microphone array recordings (p_m on $\mathcal{A}$) and HRTF datasets (h_ℓ on $\mathcal{B}$).

right ear, respectively. Each entry of $\boldsymbol{h}$ is characterized for an arbitrary position $\vec{b}_\ell$. The finite set $\{\vec{b}_\ell\}_{\ell=1,2,\ldots,L}$ is a sampling of the region of space $\mathcal{B}$ that is considered to obtain $\boldsymbol{h}$.

Finally, $\boldsymbol{C}$ is a combination matrix whose entries are acoustic transfer functions between every two points $\vec{a}_m$ and $\vec{b}_\ell$. The entries of this matrix need to be measured or calculated before hand. The conditions to calculate its pseudo-inverse, denoted by $\boldsymbol{C}^+$, are investigated below for each binaural synthesis approach. A combination matrix $\boldsymbol{C}_{\text{HRTF}}$ of size $L \times M$ (from $\vec{a}_m$ to $\vec{b}_\ell$) will be used in the HRTF modeling approach, whereas a combination matrix $\boldsymbol{C}_{\text{mic}}$ of size $M \times L$ (from $\vec{b}_\ell$ to $\vec{a}_m$) will be used in the microphone signal modeling approach. The structures for spatial signal processing that result from these two approaches are illustrated in Fig. 2.

3.2.1. HRTF modeling approach

The binaural signals are synthesized as follows:

$$\varphi = \overline{\boldsymbol{w}^\top} \boldsymbol{p}, \tag{16}$$

where

$$\boldsymbol{w} = \boldsymbol{C}^+_{\text{HRTF}} \boldsymbol{h} \tag{17}$$

defines the weighting filters organized in the matrix

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{w}^{\text{left}} \\ \boldsymbol{w}^{\text{right}} \end{bmatrix}^\top = \begin{bmatrix} w_1^{\text{left}} & w_2^{\text{left}} & \cdots & w_M^{\text{left}} \\ w_1^{\text{right}} & w_2^{\text{right}} & \cdots & w_M^{\text{right}} \end{bmatrix}^\top. \tag{18}$$

It can be verified that equating (17) and (16) results in an expression that is similar to (12) if

$$\boldsymbol{C}^+ = \overline{\boldsymbol{C}^+_{\text{HRTF}}}^\top. \tag{19}$$

Equation (17) is the result of solving a linear system that approximates the HRTF dataset according to

$$\boldsymbol{C}_{\text{HRTF}} \boldsymbol{w} = \boldsymbol{h} + \epsilon_{\text{HRTF}}, \tag{20}$$

where ϵ_{HRTF} is the approximation error. A solution to (20) can be found in terms of a pseudo-inverse $\boldsymbol{C}^+_{\text{HRTF}}$ (from $\vec{b}_\ell$ to $\vec{a}_m$) that fulfills the following condition:

$$\boldsymbol{C}^+_{\text{HRTF}} \boldsymbol{C}_{\text{HRTF}} = \boldsymbol{I}_M, \tag{21}$$

where $\boldsymbol{I}_M$ is the identity matrix of size M. This property can be regarded as a discrete version of (6).

3.2.2. Microphone signal modeling approach

The binaural signals in this case are synthesized by

$$\varphi = \overline{\boldsymbol{h}^\top} \boldsymbol{u}, \tag{22}$$

where

$$\boldsymbol{u} = \boldsymbol{C}^+_{\text{mic}} \boldsymbol{p} \tag{23}$$

defines the driving signals organized in the vector

$$\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_L \end{bmatrix}^\top. \tag{24}$$

It can be verified that replacing (23) in (22) results in an equation similar to (12) if

$$\boldsymbol{C}^+ = \boldsymbol{C}^+_{\text{mic}}. \tag{25}$$

Equation (23) arises from solving a linear system that approximates the microphone array signals in such a way that

$$\boldsymbol{C}_{\text{mic}} \boldsymbol{u} = \boldsymbol{p} + \epsilon_{\text{mic}}, \tag{26}$$

where ϵ_{mic} is the approximation error. A solution to (26) can be found in terms of a pseudo-inverse $\boldsymbol{C}^+_{\text{mic}}$ (from $\vec{a}_m$ to $\vec{b}_\ell$) that fulfills the following condition:

$$\boldsymbol{C}^+_{\text{mic}} \boldsymbol{C}_{\text{mic}} = \boldsymbol{I}_L, \tag{27}$$

where $\boldsymbol{I}_L$ is the identity matrix of size L. This condition is analogous to a discretization of (11).

If the combination matrices $\boldsymbol{C}_{\text{HRTF}}$ and $\boldsymbol{C}_{\text{mic}}$ are obtained under the same physical and geometrical conditions, and for the same number of microphones (M) and HRTFs (L), it is verified then from (19) and (25) that one matrix results to be the conjugate transpose of the other, that is,

$$\boldsymbol{C}_{\text{HRTF}} = \overline{\boldsymbol{C}_{\text{mic}}}^\top. \tag{28}$$

A remarkable relation between the two synthesis approaches is further verified when comparing the dimensions of the corresponding linear systems in (20) and (26). When one approach arises from the solution to an overdetermined system of equations, the other corresponds to an underdetermined system of equations, and vice versa.

4. PSEUDO-INVERSION BY DIAGONALIZATION

The calculation of $\boldsymbol{C}^{+}_{\mathrm{HRTF}}$ or $\boldsymbol{C}^{+}_{\mathrm{mic}}$ is simplified when the corresponding combination matrices $\boldsymbol{C}_{\mathrm{HRTF}}$ or $\boldsymbol{C}_{\mathrm{mic}}$ have representations that decouple the positions of microphones and HRTFs. In general, this is possible when a combination matrix is factored into a diagonal form, which exists by virtue of the spectral theorem [73]. A diagonal form is equivalent to having two representations of a matrix, one along its rows and the other along its columns, between which a one-to-one relation can be established by means of a diagonal matrix.

Singular value decomposition (SVD) [74] is among the most used diagonalization methods. In general, a combination matrix $\boldsymbol{C}$ can be factored in such a way that

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}\overline{\boldsymbol{V}}^{\top}. \tag{29}$$

Here, $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices, whereas $\boldsymbol{\Sigma}$ is a rectangular diagonal matrix containing the singular values σ_i, where $i = 1, \cdots, I$, and I is equal to L or M depending on whether HRTF modeling or microphone signal modeling is respectively used. The singular values are further arranged in such a way that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_I$.

The pseudo-inverse of $\boldsymbol{C}$ is calculated as follows:

$$\boldsymbol{C}^{+} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{U}}^{\top}. \tag{30}$$

The entries of $\boldsymbol{\Sigma}^{-1}$ are simply given by σ_i^{-1}. There exists an equivalence between (30) and the Moore–Penrose pseudo-inverse, $\boldsymbol{C}^{+} = [\overline{\boldsymbol{C}}^{\top}\boldsymbol{C}]^{-1}\overline{\boldsymbol{C}}^{\top}$, which is obtained by minimizing the approximation error (ϵ_{HRTF} or ϵ_{mic}) in the least-squares sense [74]. To deal with ill-conditioned matrices, inverse singular values σ_i^{-1} below a specified threshold are simply discarded. This procedure defines the so-called truncated SVD.

A more robust way of calculating a pseudo-inverse relies on applying a smooth damping to the inverse singular values in $\boldsymbol{\Sigma}^{-1}$ instead of simply discarding them by truncation. This is possible by specifying a parameter λ whose action is [75]:

$$\boldsymbol{C}^{+} = \boldsymbol{V}\left[\mathrm{diag}\left(\frac{|\sigma_i|^2}{|\sigma_i|^2 + \lambda^2}\right)\boldsymbol{\Sigma}^{-1}\right]\overline{\boldsymbol{U}}^{\top}. \tag{31}$$

An equivalence between (31) and (30) is clear for $\lambda = 0$. For $\lambda > 0$, (31) is equivalent to using the Tikhonov regularization method [75] to calculate a pseudo-inverse as follows: $\boldsymbol{C}^{+} = [\overline{\boldsymbol{C}}^{\top}\boldsymbol{C} + \lambda^2\boldsymbol{I}]^{-1}\overline{\boldsymbol{C}}^{\top}$. The parameter λ is thus known as the regularization parameter.

No consideration regarding specific geometries that underlie the description of combination matrices has been taken into account hitherto. This was a convenient pathway towards the formulation of binaural synthesis, as the resulting structures are applicable in arbitrary geometries, as long as measuring acoustic transfer functions comprising the combination matrices is possible. However, regarding the time that would demand a complete survey of the geometrical and physical conditions of the binaural system, using mathematical formulas to characterize the combination matrices is usually convenient, as long as analytic expressions describing the underlying physical phenomena exist and can be used in the modeling of combination matrices. In this regard, it is known from the literature that acoustic propagation models based on the solutions to the wave equation exist for the eleven systems of Stackel that allows for the separation of spatial coordinates [76], and also for another set of coordinate systems constructed by conformal mapping [77].

In the next two sections, the separable spherical coordinate system is used to describe a physical model for combination matrices, which is then used to present a simple but illustrative example of binaural synthesis from spherical arrays.

5. COMBINATION MATRICES FOR SPHERICAL ARRAYS

The capture of sound with uniform resolution along directions is possible by using a spherical array of microphones for recording [78,79] and a spherical array of sources for characterizing the HRTF dataset [27–29]. The use of an acoustically rigid spherical baffle for recording, where the microphone array is mounted, is particularly convenient because it adds stability during pseudo-inversion, as opposed to the so-called open microphone arrays that do not use a rigid baffle [78,79].

Spherical arrays further enable the use of modal representations of microphone recordings and HRTF datasets in terms of solutions to the acoustic wave equation at different directional resolutions [80]. These modal representations enjoy popularity because they provide an encoding and decoding scheme of directional sound information at scalable resolutions [33,34,45–55]. A directional resolution at a specific scale can be associated with a single parameter that is called the order, which is denoted by ν in this section.

In the spherical coordinate system shown in Fig. 3, a point in space $\vec{r} = (r, \theta, \phi)$ is specified by its radial distance r, azimuthal angle $\theta \in [-\pi, \pi]$, and elevation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Angles are merged into a single variable $\Omega = (\theta, \phi)$ in such a way that a point in space is also represented by $\vec{r} = (r, \Omega)$.

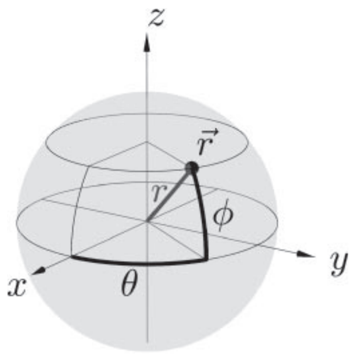


Fig. 3 Spherical coordinate system. The origin coincides with the center of the array and the center of the listener's head.

When a spherical microphone array of radius a is used for recording, each recording signal p_m of $\boldsymbol{p}$ in (14) corresponds to a microphone position $\vec{a}_m = (a, \Omega_m)$. The presence of a rigid spherical baffle of radius a is further assumed. On the other hand, when an open spherical array of radius b is used to characterize the HRTFs, each entry h_ℓ^{left} or h_ℓ^{right} of the HRTF dataset $\boldsymbol{h}$ in (15) corresponds to a position $\vec{b}_\ell = (b, \Omega_\ell)$. The use of an open spherical array in this example aims to simulate the typical way of characterizing HRTF datasets in anechoic conditions [27–29].

Analytical expressions to calculate $\boldsymbol{C}^+_{\text{HRTF}}$ or $\boldsymbol{C}^+_{\text{mic}}$ for the spherical arrays described above can be derived by first representing the entries of $\boldsymbol{C}_{\text{HRTF}}$ or $\boldsymbol{C}_{\text{mic}}$ in terms of the acoustic scattering model for a rigid sphere [80], and by subsequently performing a pseudo-inversion based on analytic Tikhonov regularization [75]. The expressions are detailed below.

In the HRTF modeling approach, the entries of $\boldsymbol{C}^+_{\text{HRTF}} = [c^{\text{HRTF}+}_{m\ell}]$ to be used in (17) are defined by

$$c^{\text{HRTF}+}_{m\ell} = \frac{\exp(-jkb)}{b}\alpha_m\beta_\ell \times \sum_{\nu=0}^{\lfloor\sqrt{L}\rfloor-1}(2\nu+1)\overline{R_\nu^{\text{reg}}(a,b,k)}P_\nu(\cos\Theta_{m\ell}). \tag{32}$$

The angular part of the sum in (32) is defined by the Legendre polynomial P_ν of order ν evaluated at the cosine of the angle $\Theta_{\ell m}$ between $\vec{b}_\ell$ and $\vec{a}_m$. The radial part of the sum in (32) is defined by the regularized radial filter

$$R_\nu^{\text{reg}} = \frac{|R_\nu|^2}{|R_\nu|^2+\lambda^2}\times\frac{1}{R_\nu}, \quad \text{where} \quad R_\nu = \frac{h_\nu(kb)}{ka^2 h'_\nu(ka)}. \tag{33}$$

Here, λ is the regularization parameter, h_ν is the spherical Hankel function of the second kind and order ν, and $'$ indicates the derivative with respect to the argument. Finally, α_m and β_ℓ in (32) respectively denote discrete versions of the infinitesimals $d\vec{a}$ and $d\vec{b}$ in (1). These integration quadratures are positive, normalized so that $\sum_m \alpha_m = \sum_\ell \beta_\ell = 1$, and depend on the spherical sampling schemes [81] used to decide the positions of microphones and HRTFs.

In the microphone signal modeling approach, the entries of $\boldsymbol{C}^+_{\text{mic}} = [c^{\text{mic}+}_{\ell m}]$ to be used in (23) are defined by

$$c^{\text{mic}+}_{\ell m} = \frac{\exp(jkb)}{b}\alpha_m\beta_\ell \times \sum_{\nu=0}^{\lfloor\sqrt{M}\rfloor-1}(2\nu+1)R_\nu^{\text{reg}}(a,b,k)P_\nu(\cos\Theta_{m\ell}). \tag{34}$$

As in (32), the angular part of the sum in (34) is defined by the Legendre polynomial P_ν, the radial part is defined by (33), and the normalized integration quadratures α_m and β_ℓ depend on the sampling schemes under consideration.

A direct inspection of (31) and (33) brings to light an important relation between singular values and radial filters. In spherical geometries, $\overline{R_\nu}$ are the singular values of $\boldsymbol{C}_{\text{HRTF}}$, whereas R_ν are the singular values of $\boldsymbol{C}_{\text{mic}}$. A similar relation has also been reported in a study concerning the modeling of HRTFs with spherical harmonic functions [82].

Diagonal forms of $\boldsymbol{C}_{\text{HRTF}}$ and $\boldsymbol{C}_{\text{mic}}$ to analytically decouple the microphone directions Ω_m and the HRTF directions Ω_ℓ exist owing to the Legendre addition theorem [80]:

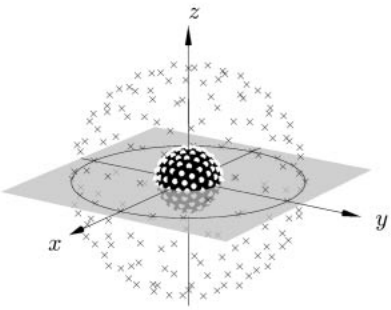
$$P_\nu(\cos\Theta_{m\ell}) = \frac{4\pi}{2\nu+1}\sum_{\mu=-\nu}^{\nu} Y_\nu^\mu(\Omega_m)\overline{Y_\nu^\mu(\Omega_\ell)}, \tag{35}$$

where Y_ν^μ are the so-called spherical harmonic functions of order ν and degree μ.

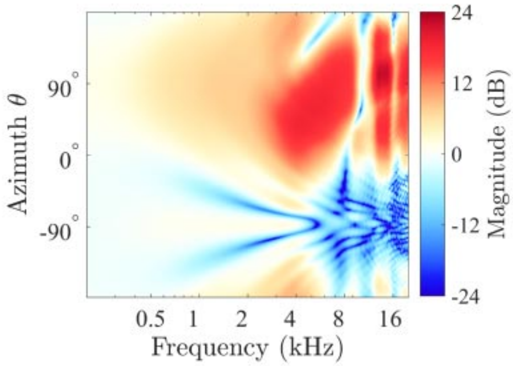
The Legendre addition theorem in (35) provides the theoretical foundations for two popular binaural synthesis methods in spherical geometries. The first method relies on a diagonalization of $\boldsymbol{C}_{\text{HRTF}}$ and is known as binaural modal beamforming [34,52,53]. The second method relies on a diagonalization of $\boldsymbol{C}_{\text{mic}}$ and is known as binaural ambisonics [33,48,50]. In both cases, the pseudo-inverses obtained by analytic diagonalization can be formulated as follows:

$$\boldsymbol{C}^+ = \boldsymbol{D}\left[\operatorname{diag}\left(\frac{|R_\nu|^2}{|R_\nu|^2+\lambda^2}R_\nu^{-1}\right)\right]\boldsymbol{E}^+. \tag{36}$$

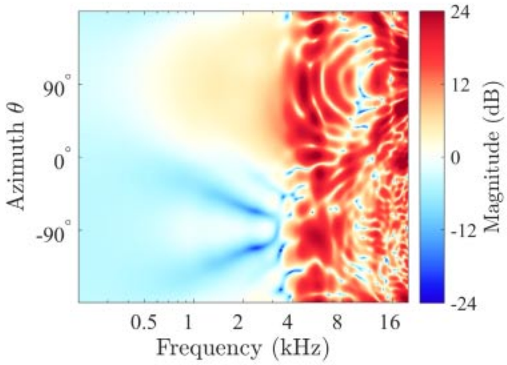
Here, $\boldsymbol{E}^+$ is the directional encoding matrix, and $\boldsymbol{D}$ is the directional decoding matrix. Both matrices are defined in terms of spherical harmonic functions. $\boldsymbol{E}^+$ is equivalent to a spherical Fourier transform, whereas $\boldsymbol{D}$ is equivalent to an inverse spherical Fourier transform. In spherical geometries, the mapping to the singular value domain remains described by analytic directional representations based on spherical Fourier transforms. A complete analogy can thus be established between the unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in (31) and the Fourier transform matrices in (36). Furthermore, the regularized singular values of $\boldsymbol{C}_{\text{HRTF}}$ are equal to



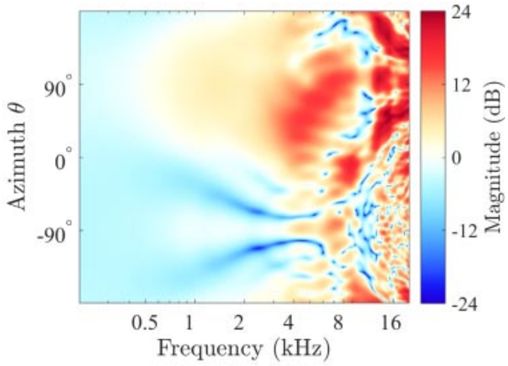
(a) Geometry.



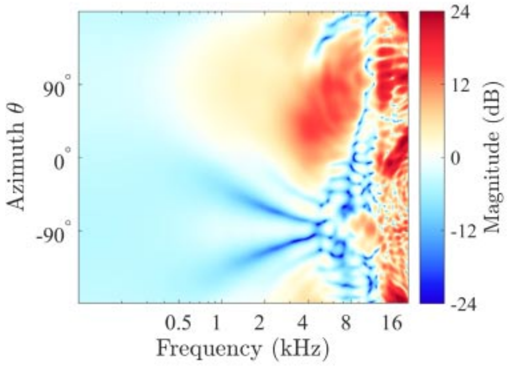
(b) Reference binaural signals.



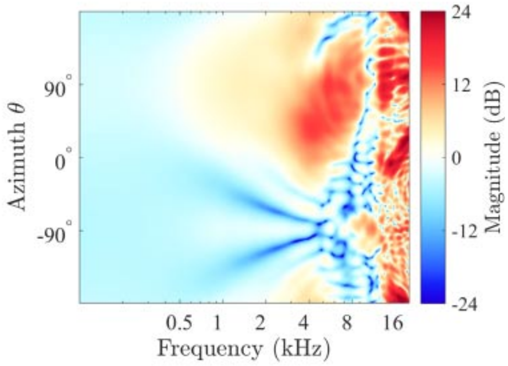
(c) HRTF modeling, $M = 92, L = 252$.



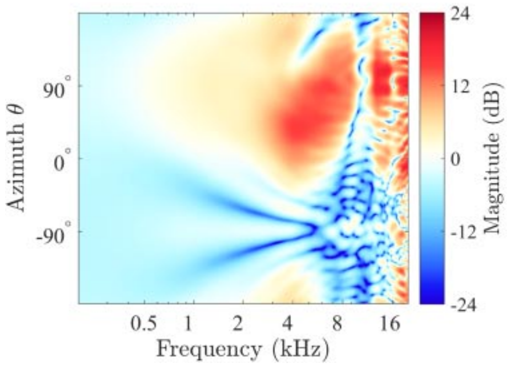
(d) Microphone signal modeling, $M = 92, L = 252$.



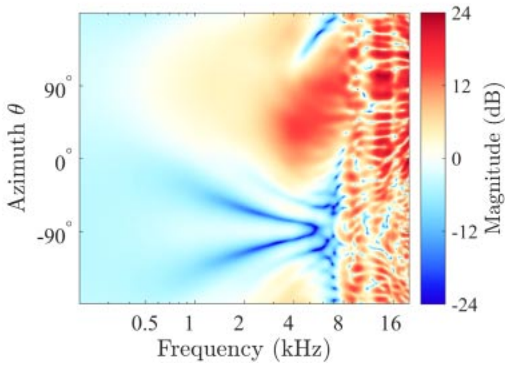
(e) HRTF modeling, $M = L = 252$.



(f) Microphone signal modeling, $M = L = 252$.



(g) HRTF modeling, $M = 492, L = 252$.



(h) Microphone signal modeling, $M = 492, L = 252$.

Fig. 4 Binaural synthesis examples.

the regularized radial filters $\overline{R_\nu^{\text{reg}}}$, whereas the regularized singular values of $\boldsymbol{C}_{\text{mic}}$ are equal to the regularized radial filters R_ν^{reg} in (33).

6. EXAMPLES OF BINAURAL SYNTHESIS

Synthesis examples for distinct numbers of microphones (M) and HRTFs (L) are presented in this section. The radii of the arrays are $a = 8.5$ cm and $b = 1.5$ m. The positioning of microphones and HRTFs was decided according to spherical grids constructed by subdividing the edges of an icosahedron [83], such as the icosahedral grids used in the SENZI system [35–39]. Examples of such icosahedral grids are shown in Fig. 4(a), where white dots and black marks respectively indicate microphones and HRTFs.

The microphone signals were simulated with the algorithm described in [84], whereas the HRTF datasets were calculated using the boundary element method solver described in [20]. Synthesis examples were restricted to target sound sources lying on the horizontal black circle in Fig. 4(a), at the same distance used to obtain the HRTF dataset. The reference set of binaural signals for such target sources is shown in Fig. 4(b). The synthesized binaural signals are shown in Figs. 4(c)–4(h). The regularization parameter to calculate the pseudo-inverses was set to $\lambda = 1 \times 10^{-3}$.

When $M = L$, $\boldsymbol{C}_{\text{HRTF}}$ and $\boldsymbol{C}_{\text{mic}}$ are equivalent because they both correspond to a determined linear system. Figures 4(e) and 4(f) show that both approaches yield the same synthesis results in the determined case.

When $M < L$, $\boldsymbol{C}_{\text{HRTF}}$ represents an overdetermined system, whereas $\boldsymbol{C}_{\text{mic}}$ represents an underdetermined system. A comparison between Figs. 4(c) and 4(d) show that more distortion at higher frequencies is observed in the overdetermined case in Fig. 4(c).

Finally, when $M > L$, $\boldsymbol{C}_{\text{HRTF}}$ corresponds to an underdetermined system, whereas $\boldsymbol{C}_{\text{mic}}$ corresponds to an overdetermined system. A comparison between Figs. 4(g) and 4(h) show again that more distortion at higher frequencies is observed in the overdetermined case in Fig. 4(h).

7. SUMMARY

A review of signal processing methods to accurately synthesize the sound pressure at the ears from microphone array recordings was presented. Special attention was given to a modern class of methods that rely on combining the spatial information available in microphone array recordings with the spatial information available in datasets of head-related transfer functions (HRTFs). The use of a microphone array offers the possibility of including the dynamic auditory cues used in sound localization because the movements of listeners can be tracked through digital computations in real-time or non-real-time conditions. The HRTF datasets allow the inclusion of the auditory localization cues that correspond to the individual traits of the external anatomy of the listeners.

Table 2 Two formulations of binaural synthesis.

Characteristic	Approach: HRTF modeling	Approach: Microphone signal modeling
Linear system formulation	Acoustic transfer functions in $\boldsymbol{C}_{\text{HRTF}}$ are linearly combined to approximate the HRTF dataset $\boldsymbol{h}$: $\boldsymbol{C}_{\text{HRTF}}\boldsymbol{w} = \boldsymbol{h} + \epsilon_{\text{HRTF}}$.	Acoustic transfer functions in $\boldsymbol{C}_{\text{mic}}$ are linearly combined to approximate the microphone signals $\boldsymbol{p}$: $\boldsymbol{C}_{\text{mic}}\boldsymbol{u} = \boldsymbol{p} + \epsilon_{\text{mic}}$.
Synthesis algorithm	(1) Weighting filters $\boldsymbol{w} = \boldsymbol{C}_{\text{HRTF}}^{+}\boldsymbol{h}$. (2) Binaural signals $\varphi = \overline{\boldsymbol{w}}^{\top}\boldsymbol{p}$.	(1) Driving signals $\boldsymbol{u} = \boldsymbol{C}_{\text{mic}}^{+}\boldsymbol{p}$. (2) Binaural signals $\varphi = \overline{\boldsymbol{h}}^{\top}\boldsymbol{u}$.

A general formulation for this modern class of methods was presented in terms of a linear system of equations, whose associated matrix is composed of acoustic transfer functions that relate the positions of microphones and HRTFs. Based on this formulation, it was shown that most of the existing methods under consideration can be classified into two prominent approaches: 1) the HRTF modeling approach and 2) the microphone signal modeling approach. An important relation between these two approaches was evidenced in the general formulation: when one approach arises from the solution to an overdetermined system, the other corresponds to an underdetermined system, and vice versa. Simulation examples in spherical geometries were also presented to illustrate the fact that underdetermined systems generally outperform overdetermined ones. The formulation of the underlying linear systems and their corresponding synthesis algorithms are summarized in Table 2.

Other kinds of spatial information can further be included as parameters in the combination matrices $\boldsymbol{C}_{\text{HRTF}}$ or $\boldsymbol{C}_{\text{mic}}$. For instance, spatial patterns for artificial reverberation or directional selectivity can further be included when the targeted application does not aim to re-create an auditory scene but instead requires additional spatial edition capabilities. Such is the case, for example, of binaural systems that use an additional beamforming stage to increase the intelligibility of speech [52,53]. In this regard, recent trends in array signal processing can be exploited to optimally represent $\boldsymbol{C}_{\text{HRTF}}$ or $\boldsymbol{C}_{\text{mic}}$ for future recording, editing, and reproduction of sound in binaural systems. Compressive sensing methods [85] that reduce the sparseness of array signals can for instance be used to optimize the directivity of microphone arrays without affecting the spatial patterns of the HRTF datasets.

ACKNOWLEDGMENTS

A part of this study was supported by a JSPS Grant-in-Aid for Scientific Research (nos. JP24240016 and JP16H01736) and the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet." We thank Makoto Otani for allowing us to use his BEM solver to calculate HRTF datasets.

REFERENCES

[1] H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.*, **36**, 171–218 (1992).

[2] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn.* (*E*), **1**, 167–174 (1980).

[3] V. Algazi and R. Duda, "Headphone-based spatial sound," *IEEE Signal Process. Mag.*, **28**, 33–42 (2011).

[4] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, **37**, 3–19 (1989).

[5] T. Betlehem, W. Zhang, M. Poletti and T. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, **32**, 81–91 (2015).

[6] H. Wallach, "On sound localization," *J. Acoust. Soc. Am.*, **10**, 270–274 (1939).

[7] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *J. Acoust. Soc. Am.*, **42**, 480–488 (1967).

[8] J. Kawaura, Y. Suzuki, F. Asano and T. Sone, "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," *J. Acoust. Soc. Jpn.* (*J*), **45**, 756–766 (1989) (in Japanese), English translation: *J. Acoust. Soc. Jpn.* (*E*), **12**, 203–216 (1991).

[9] 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).

[10] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.*, **44**, 240–249 (1968).

[11] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. (MIT Press, Cambridge, MA, London, 1997).

[12] N. I. Durlach, A. Rigopulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn and E. M. Wenzel, "On the externalization of auditory images," *Presence-Teleop. Virt.*, **1**, 251–257 (1992).

[13] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, **99**, 3678–3688 (1996).

[14] I. Toshima, H. Uematsu and T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement: TeleHead," *Acoust. Sci. & Tech.*, **24**, 327–329 (2003).

[15] I. Toshima, S. Aoki and T. Hirahara, "Sound localization using an acoustical telepresence robot: TeleHead II," *Presence-Teleop. Virt.*, **17**, 392–404 (2008).

[16] V. R. Algazi, R. O. Duda and D. M. Thompson, "Motion-tracked binaural sound," *J. Audio Eng. Soc.*, **52**, 1142–1156 (2004).

[17] D. N. Zotkin, R. Duraiswami, E. Grassi and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *J. Acoust. Soc. Am.*, **120**, 2202–2215 (2006).

[18] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.*, **110**, 2440–2448 (2001).

[19] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.*, **110**, 2449–2455 (2001).

[20] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, **119**, 2589–2598 (2006).

[21] P. Mokhtari, H. Takemoto, R. Nishimura and H. Kato, "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data," in *AES 123rd Convention*, Oct. (2007).

[22] K. Matsui and A. Ando, "Estimation of individualized head-related transfer function based on principal component analysis," *Acoust. Sci. & Tech.*, **30**, 338–347 (2009).

[23] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami and D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *J. Acoust. Soc. Am.*, **127**, 370–386 (2010).

[24] A. Meshram, R. Mehra and D. Manocha, "Efficient HRTF computation using adaptive rectangular decomposition," in *AES 55th Int. Conf. Spatial Audio*, Aug. (2014).

[25] H. Ziegelwanger, P. Majdak and W. Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *J. Acoust. Soc. Am.*, **138**, 208–222 (2015).

[26] V. Algazi, R. Duda, D. Thompson and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE WASPAA*, pp. 99–102 (2001). [Online] Available: http://interface.cipic.ucdavis.edu/sound/hrtf.html

[27] S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki and T. Sone, "A database of head-related transfer functions in whole directions on upper hemisphere," *Acoust. Sci. & Tech.*, **23**, 160–162 (2002).

[28] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger and M. Noisternig, "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," in *AES 134th Convention*, May (2013). [Online] Available: https://www.sofaconventions.org/

[29] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane and S. Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoust. Sci. & Tech.*, **35**, 159–165 (2014). [Online] Available: http://www.riec.tohoku.ac.jp/pub/hrtf/index.html

[30] S. Takane, Y. Suzuki and T. Sone, "A new method for global sound field reproduction based on Kirchhoff's integral equation," *Acta Acust. united Ac.*, **85**, 250–257 (1999).

[31] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, "ADVISE: A new method for high definition virtual acoustic display," in *Proc. Int. Conf. Auditory Display*, Jul. (2002).

[32] 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).

[33] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov and L. S. Davis, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *AES 119th Convention*, New York, USA, Oct. (2005).

[34] W. Song, W. Ellermeier and J. Hald, "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *J. Acoust. Soc. Am.*, **123**, 910–924 (2008).

[35] S. Sakamoto, S. Hongo, R. Kadoi and Y. Suzuki, "SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," in *Proc. 2nd Int. Symp. Universal Commun.*, pp. 429–434 (2008).

[36] S. Sakamoto, J. Kodama, S. Hongo, T. Okamoto, Y. Iwaya

and Y. Suzuki, “A 3D sound-space recording system using spherical microphone array with 252ch microphones,” in *Proc. 20th Int. Congr. Acoust.*, Sydney, Australia, Aug. (2010).

[37] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya and Y. Suzuki, “Improvement of accuracy of three-dimensional sound space synthesized by real-time “SENZI,” a sound space information acquisition system using spherical array with numerous microphones,” *Proc. Mtgs. Acoust.*, Vol. 19, No. 1 (2013).

[38] S. Sakamoto, S. Hongo and Y. Suzuki, “3D sound-space sensing method based on numerous symmetrically arranged microphones,” *IEICE Trans. Fundam.*, **E97-A**, 1893–1901 (2014).

[39] 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).

[40] M. V. Laitinen and V. Pulkki, “Binaural reproduction for directional audio coding,” in *Proc. IEEE WASPAA*, Espoo, Finland, Oct., pp. 337–340 (2009).

[41] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. van de Par, V. Mellert and D. Püschel, “Robustness of virtual artificial head topologies with respect to microphone positioning errors,” in *Proc. Forum Acusticum*, Aalborg, pp. 2251–2256 (2011).

[42] E. Rasumow, M. Blau, S. Doclo, M. Hansen, S. Van de Par, D. Püschel and V. Mellert, “Least squares versus non-linear cost functions for a vitual artificial head,” *Proc. Mtgs. Acoust.*, Vol. 19, No. 1 (2013).

[43] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Doclo, V. Mellert and D. Püschel, “Smoothing individual head-related transfer functions in the frequency and spatial domains,” *J. Acoust. Soc. Am.*, **135**, 2012–2025 (2014).

[44] E. Rasumow, M. Hansen, S. v. d. Par, D. Püschel, V. Mellert, S. Doclo and M. Blau, “Regularization approaches for synthesizing HRTF directivity patterns,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **24**, 215–225 (2016).

[45] S. Spors, H. Wierstorf and M. Geier, “Comparison of modal versus delay-and-sum beamforming in the context of data-based binaural synthesis,” in *AES 132nd Convention*, Budapest, Hungary, Apr. (2012).

[46] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf and B. Rafaely, “Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution,” *J. Acoust. Soc. Am.*, **133**, 2711–2721 (2013).

[47] C. D. Salvador, S. Sakamoto, J. Treviño, J. Li, Y. Yan and Y. Suzuki, “Accuracy of head-related transfer functions synthesized with spherical microphone arrays,” *Proc. Mtgs. Acoust.*, Vol. 19, No. 1, Apr. (2013).

[48] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Numerical evaluation of binaural synthesis from rigid spherical microphone array recordings,” in *AES Int. Conf. Headphone Technology*, Aalborg, Denmark, Aug. (2016).

[49] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Evaluation of white noise gain in a binaural system for microphone arrays,” in *Proc. Autumn Meet. Acoust. Soc. Jpn.*, Toyama, pp. 401–404 (2016).

[50] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Spatial accuracy of binaural synthesis from rigid spherical microphone array recordings,” *Acoust. Sci. & Tech.*, **38**, 23–30 (2017).

[51] B. Bernschütz, A. V. Giner, C. Pörschmann and J. Arend, “Binaural reproduction of plane waves with reduced modal order,” *Acta Acust. united Ac.*, **100**, 972–983 (2014).

[52] N. R. Shabtai and B. Rafaely, “Generalized spherical array beamforming for binaural speech reproduction,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **22**, 238–247 (2014).

[53] N. R. Shabtai, “Optimization of the directivity in binaural sound reproduction beamforming,” *J. Acoust. Soc. Am.*, **138**, 3118–3128 (2015).

[54] J. Sheaffer, S. Villeval and B. Rafaely, “Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction,” in *Proc. EAA Jt. Symp. Auralization and Ambisonics*, Berlin, Germany, Mar. (2014).

[55] J. Sheaffer, M. van Walstijn, B. Rafaely and K. Kowalczyk, “Binaural reproduction of finite difference simulations using spherical array processing,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **23**, 2125–2135 (2015).

[56] S. Delikaris-Manias, J. Vilkamo and V. Pulkki, “Parametric binaural rendering utilizing compact microphone arrays,” in *Proc. IEEE ICASSP*, Apr., pp. 629–633 (2015).

[57] R. D. Shilling and B. Shinn-Cunningham, “Virtual auditory displays,” in *Handbook of Virtual Environment Technology* (Lawrence Erlbaum Associates, New York, 2002), pp. 65–92.

[58] S. Tachi, M. Sato and M. Hirose, *Virtual Reality* (Corona Publishing Co., Ltd., Tokyo, 2010).

[59] M. J. Evans, “Analyzing head-related transfer functions measurements using surface spherical harmonics,” *J. Acoust. Soc. Am.*, **104**, 2400–2411 (1998).

[60] K. Watanabe, S. Takane and Y. Suzuki, “Interpolation of head-related transfer functions based on the common-acoustical-pole and residue model,” *Acoust. Sci. & Tech.*, **24**, 335–337 (2003).

[61] K. Watanabe, S. Takane and Y. Suzuki, “A novel interpolation method of HRTFs based on the common-acoustical-pole and zero model,” *Acta Acust. united Ac.*, **91**, 958–966 (2005).

[62] M. Ohuchi, Y. Iwaya, Y. Suzuki and T. Munekata, “Software engine for multipurpose virtual auditory displays and its application in training systems for sound space perception,” *J. Acoust. Soc. Jpn.* (*J*), **62**, 224–232 (2006) (in Japanese).

[63] T. Ajdler, C. Faller, L. Sbaiz and M. Vetterli, “Sound field analysis along a circle and its applications to HRTF interpolation,” *J. Audio Eng. Soc.*, **56**, 156–175 (2008).

[64] A. Kan, C. Jin and A. van Schaik, “A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function,” *J. Acoust. Soc. Am.*, **125**, 2233–2242 (2009).

[65] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer and M. Noisternig, “Calculation of head-related transfer functions for arbitrary field points using spherical harmonics,” *Acta Acust. united Ac.*, **98**, 72–82 (2012).

[66] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “A new signal processing procedure for stable distance manipulation of circular HRTFs on the horizontal plane,” in *Proc. Spring Meet. Acoust. Soc. Jpn.*, Yokohama, pp. 561–564 (2016).

[67] C. D. Salvador, S. Sakamoto, J. Treviño and Y. Suzuki, “Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values,” *Acoust. Sci. & Tech.*, **38**, 1–13 (2017).

[68] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini and L. S. Davis, “HRTF personalization using anthropometric measurements,” in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct., pp. 157–160 (2003).

[69] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki and K. Aso, “Estimation of interaural level difference based on anthropometry and its effect on sound localization,” *J. Acoust. Soc. Am.*, **122**, 2832–2841 (2007).

[70] S. Xu, Z. Li and G. Salvendy, “Individualization of head-related transfer function for three-dimensional virtual auditory display: A review,” in *Proc. 2nd Int. Conf. Virtual Reality*, R. Shumaker, Ed. (Springer, Berlin, Heidelberg, 2007), pp. 397–

407.
[71] S. Yairi, Y. Iwaya and Y. Suzuki, "Individualization feature of head-related transfer functions based on subjective evaluation," in *Proc. Int. Conf. Auditory Display* (2008).
[72] G. H. Koopmann, L. Song and J. B. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," *J. Acoust. Soc. Am.*, **86**, 2433–2438 (1989).
[73] P. R. Halmos, "What does the spectral theorem say?" *Am. Math. Mon.*, **70**, 241–247 (1963).
[74] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, **14**, 403–420 (1970).
[75] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Rev.*, **40**, 636–666 (1998).
[76] L. P. Eisenhart, "Separable systems of stackel," *Ann. Math.*, **35**, 284–305 (1934).
[77] P. Moon and D. E. Spencer, *Field Theory Handbook: Including Coordinate Systems, Differential Equations and Their Solutions* (Springer, Berlin, 2012).
[78] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE ICASSP*, vol. II, Orlando, FL, USA, May, pp. 1781–1784 (2002).
[79] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, **13**, 135–143 (2005).
[80] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, 1999).
[81] I. H. Sloan and R. S. Womersley, "Extremal systems of points and numerical integration on the sphere," *Adv. Comput. Math.*, **21**, 107–125 (2003).
[82] P. A. Nelson and Y. Kahana, "Spherical harmonics, singular-value decomposition and the head-related transfer function," *J. Sound Vib.*, **239**, 607–637 (2001).
[83] R. Sadourny, A. Arakawa and Y. Mintz, "Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere 1," *Mon. Wea. Rev.*, **96**, 351–356 (1968).
[84] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, **104**, 3048–3058 (1998).
[85] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, **25**, 21–30 (2008).

César D. Salvador graduated from Pontifical Catholic University of Peru in 2005, and received his M.Sc. and Ph.D. degrees in information sciences from Tohoku University respectively in 2013 and 2016. He is currently a Postdoctoral Researcher at the Research Institute of Electrical Communication, Tohoku University. His research interests include theoretical acoustics, spatial hearing, and array signal processing.

Shuichi Sakamoto received his B.S., M.Sc., and Ph.D. degrees from Tohoku University, in 1995, 1997, and 2004, respectively. He is currently an Associate 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 the development of high-definition 3D audio recording systems. He is a Member of ASJ, IEICE, VRSJ, and other professional societies.

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 is currently an Assistant Professor in the Research Institute of Electrical Communication of Tohoku University. His research interests include sound field recording and reproduction, array signal processing, and spatial audio.

Yôiti Suzuki graduated from Tohoku University in 1976 and received his Ph.D. degree in electrical and communication engineering in 1981. He is currently a Professor at the Research Institute of Electrical Communication, Tohoku University. His research interests include psychoacoustics, multimodal perception, high-definition 3D auditory displays, and digital signal processing of acoustic signals. He has received the Awaya Kiyoshi Award and Sato Prize from the Acoustical Society of Japan, as well as the FIT Funai Best Paper Award.