

ACOUSTICAL LETTER

Mixed-order Ambisonics encoding of cylindrical microphone array signals

Jorge Trevino[1,2,*], Shoichi Koyama[3,†], 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,
6–3–09 Aramaki Aza Aoba, Aoba-ku, Sendai, 980–8577 Japan
[3]NTT Media Intelligence Laboratories,
3–9–11 Midori-cho, Musashino, 180–8585 Japan

(Received 29 November 2013, Accepted for publication 18 December 2013)

Keywords: Mixed-order Ambisonics, Cylindrical array, Spherical harmonics, Sound field reproduction
PACS number: 43.60.Fg [doi:10.1250/ast.35.174]

1. Introduction

Humans can determine the approximate position of a sound source by listening to it [1]. Systems intended to realistically present auditory information must convey the spatial features needed for its users to localize sound. A promising technique that seeks to achieve this using surrounding loudspeaker arrays is called high-order Ambisonics (HOA) [2].

HOA is a sound field reproduction method; that is, it attempts to re-create the sound pressure at all points in the neighborhood of the listener. An important feature of HOA is that it defines a system-agnostic encoding of sound fields. Sound fields recorded using any suitable system can be encoded using HOA and later reproduced using a variety of sound presentation technologies with an appropriate HOA decoder.

Sound fields carry too much information for their practical reproduction. A compromise must be made between system complexity and spatial resolution. Humans exhibit greater spatial accuracy when localizing sound sources in the horizontal plane [1]. Mixed-order Ambisonics (MOA), a variation of HOA, considers this and uses a higher resolution in the horizontal plane compared to that used for other directions [3].

Sound field recording for HOA encoding typically uses spherical microphone arrays. Recording for MOA encoding, however, requires different spatial resolutions in the horizontal plane and elsewhere. To increase horizontal resolution in a spherical array, more microphones must be placed along the equator. At the same time, the radius of the sphere cannot be too large or spatial aliasing will become a problem for the non-horizontal components, where microphone counts are kept low [4].

The present research considers the use of a cylindrical microphone array to record sound fields and encode them using MOA. The use of cylindrical arrays for sound field recording is not new [5–7]. Cylindrical microphone arrays are a good choice for robust beamforming [8]. On the other hand, cylindrical array signals cannot be directly reproduced with any sound presentation system, as is the case for HOA and MOA encodings. In this paper, we present an encoder that produces conventional MOA signals from cylindrical microphone array recordings. The results of this encoding can be decoded and reproduced with any Ambisonics-capable system.

*e-mail: jorge@ais.riec.tohoku.ac.jp
†e-mail: koyama.shoichi@lab.ntt.co.jp
‡e-mail: saka@ais.riec.tohoku.ac.jp
§e-mail: yoh@riec.tohoku.ac.jp

2. Mixed-order Ambisonics

The most salient feature of HOA is that it defines a scalable and system-agnostic encoding format for sound field information. A typical HOA system samples the sound field using a spherical microphone array. The microphone signals are processed to encode the sound pressure distribution on the sphere using the spherical harmonic functions [9]. This is known as the spherical harmonic decomposition and defines the basis of HOA. It is given by the following equations:

$$B_{nm}(k) = \int_{\theta=-\pi}^{\pi} \int_{\varphi=-\pi/2}^{\pi/2} p(k,\theta,\varphi) Y_{nm}^{*}(\theta,\varphi) \sin(\varphi) d\varphi d\theta. \quad (1)$$

$$p(k,\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_{nm}(k) Y_{nm}(\theta,\varphi). \quad (2)$$

Here, k stands for the wavenumber; the azimuth and elevation angles are denoted by θ and φ, respectively. The functions Y_{nm} are the spherical harmonics of order n and degree m (this follows the conventions of HOA; however, in physics and mathematics it is common to call n the degree and m the order). Low-order spherical harmonics are presented in Fig. 1. Equation (1) shows the encoding of the sound pressure distribution p as a set of HOA channels B_{nm}. The inverse operation, shown in Eq. (2), decodes the coefficients to re-create the sound pressure over a spherical boundary. Actual implementations truncate the first sum up to a given number N referred to as the *Ambisonic order*.

Equations (1) and (2) do not have a preferred orientation. Spatial features of the sound field on the horizontal plane are encoded with the same resolution as the rest. To use separate resolutions for the horizontal and non-horizontal components of the field, MOA uses two different Ambisonic orders. A *full-*

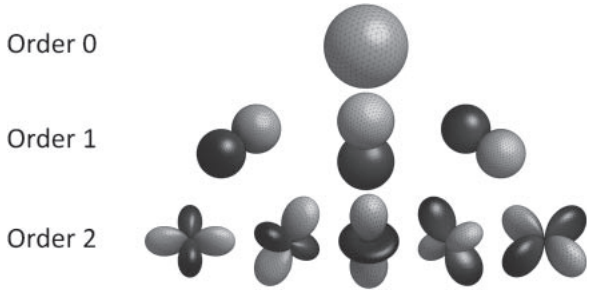

Fig. 1 Spherical harmonic functions of order 0, 1 and 2. Bright lobes indicate positive values, dark ones are negative. There are $2N+1$ spherical harmonics of order N; however, only two of them are needed to encode spatial information in the horizontal plane.

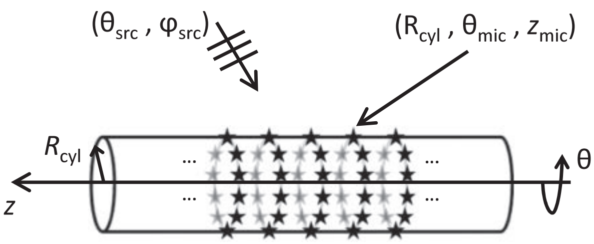

Fig. 2 Layout of a cylindrical microphone array like those used by our proposal. The microphones are arranged as equispaced parallel rings on the surface of a rigid cylinder.

3D Ambisonic order is used to encode the entire sound field by applying Eq. (1). This encoding is complemented with a high-resolution description of the sound field in the horizontal plane. This is done by continuing the spherical harmonic expansion up to a higher Ambisonic order, the *horizontal Ambisonic order*. However, only the horizontally-oriented spherical harmonics are considered. That is, only the expansion coefficients for which $m = \pm n$ are used, the rest of the coefficients are not calculated.

3. Cylindrical microphone arrays

The cylindrical microphone arrays considered in this paper consist of a set of parallel rings of equidistant microphones aligned on the surface of a rigid cylinder. This configuration is shown in Fig. 2. The cylinder's axis is assumed to lie on the horizontal plane; therefore, the number of microphones per ring will determine the spatial resolution for elevation, while the number of rings governs the horizontal resolution. The possibility of independently choosing the horizontal and polar resolutions makes cylindrical arrays an efficient choice for MOA.

The microphones are placed over a rigid cylinder to magnify the effects of the direction of arrival on the sound pressure sensed by them. This leads to a more robust recording system [8]. The sound pressure field sensed by the microphones after considering the effects of the rigid cylinder can be treated as the sum of an incident sound pressure field, ψ_i, and a scattered sound pressure field, ψ_s:

$$\psi_t(k, \boldsymbol{r}) = \psi_i(k, \boldsymbol{r}) + \psi_s(k, \boldsymbol{r}). \tag{3}$$

Vector $\boldsymbol{r}$ determines the spatial point where the total sound pressure field, ψ_t, is measured.

The objective is to encode the incident field; however, only the total field can be observed directly. We must find a relationship between these two fields in order to remove the scattered field from the final encoding. To simplify the problem, we use a result known as the plane-wave decomposition [6,7]. This technique makes it possible to treat any sound field as a superposition of plane waves arriving from all directions.

It is convenient to apply the plane-wave decomposition to the incident sound pressure field:

$$\psi_i = \boldsymbol{\Lambda} \cdot \tilde{\boldsymbol{\psi}}_{\mathbf{i}}. \tag{4}$$

The components of vector $\boldsymbol{\Lambda}$ are the weights to be applied to each of the plane waves, components of vector $\tilde{\boldsymbol{\psi}}_{\mathbf{i}}$, to recover the incident field ψ_i.

The plane waves forming vector $\tilde{\boldsymbol{\psi}}_{\mathbf{i}}$ are linearly independent and can be treated separately. The problem of removing the scattered field has been reduced to finding a relationship between each of the incident plane waves and its corresponding observed total field. Arbitrary incident and total fields will be related by the linear combination of these results.

The first step is to consider that the cylinder defines a rigid boundary; that is

$$\left.\frac{\partial}{\partial r}\psi_t(k, \boldsymbol{r})\right|_{r=R_{\mathrm{cyl}}} = 0. \tag{5}$$

Here, R_{cyl} stands for the radius of the cylinder. After introducing this constraint, the total sound field on the surface of the rigid cylinder, when the incident field consists of a plane wave arriving from an elevation angle θ_{src} and azimuth angle φ_{src}, is given by the following formula:

$$\tilde{\psi}_t^{\mathrm{src}}(k, \theta, z) = \frac{i \exp[ik \sin\varphi_{\mathrm{src}} z]}{\pi^2 k \sin\varphi_{\mathrm{src}} R_{\mathrm{cyl}}} \sum_{n=-\infty}^{\infty} \frac{i^n \exp[in(\theta - \theta_{\mathrm{src}})]}{H_n^{(1)'}(k \sin\varphi_{\mathrm{src}} R_{\mathrm{cyl}})}. \tag{6}$$

We use the notation $\tilde{\psi}_t^{\mathrm{src}}$ to emphasize this is the total field due to a plane wave incoming from a particular direction. $H_n^{(1)'}$ stands for the derivative of the Hankel function of the first kind and order n. The total sound field, in general, is given by the linear combination:

$$\psi_t = \boldsymbol{\Lambda} \cdot \tilde{\boldsymbol{\psi}}_{\mathbf{t}}. \tag{7}$$

The weights $\boldsymbol{\Lambda}$ are the same for both the incident and the total field. Calculating them will allow us to remove the scattered field.

4. Proposed system

Our proposal uses a cylindrical microphone array to record sound field information and encodes it using MOA. A block diagram for the proposed encoder is shown in Fig. 3. We introduce a *measuring grid* which does not necessarily match the microphones' layout. A set of beamformers [9] are

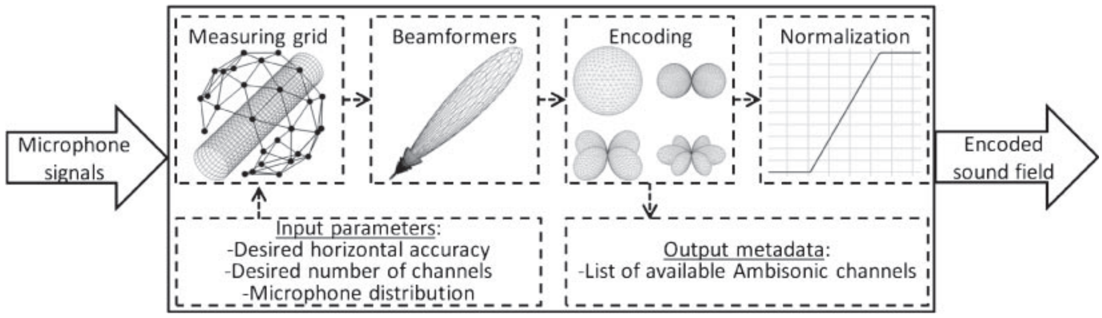

Fig. 3 Block diagram of the proposed mixed-order Ambisonics encoder for cylindrical microphone arrays.

used to isolate the sound arriving from each direction in the measuring grid. These measurements are encoded according to their angles of arrival using the spherical harmonic functions. Finally, the encoding is normalized using the Furse-Malham weighting coefficients [10] to ensure optimal use of the system's dynamic range before the signals are broadcasted or stored. The resulting encoding is in line with the conventions used in Ambisonics and can therefore be used with any system capable of reproducing MOA [2,3,11].

The horizontal accuracy of MOA encodings is set by the horizontal Ambisonic order M. In MOA, $2M+1$ channels are used to characterize the sound field in the horizontal plane. It is possible to specify a total number of channels for the full encoding with some freedom. This determines the amount of information used to encode elevation by setting the full-3D Ambisonic order N according to the following expression:

$$\text{No. channels} = N^2 + 2M + 1. \tag{8}$$

The special case of $N = M$ corresponds to conventional HOA, where both azimuth and elevation resolutions are equal. Most practical applications of MOA use relatively small values for N.

4.1. The measuring grid and beamforming

Our proposal defines a measuring grid as an almost uniform sampling of all directions. The density of the sampling is given by the constraint:

$$\text{No. directions} \geq (M+1)^2. \tag{9}$$

It is not possible to define a uniform measuring grid for an arbitrary number of directions. Nevertheless, a minimum-energy sampling of the sphere [12] can be used as a basis for the measuring grid. An illustration of this is shown in Fig. 4 with some points on the sphere omitted for clarity.

To isolate the sound field arriving from some direction $(\theta_{\text{dir}}, \varphi_{\text{dir}})$ in the measuring grid, our proposal uses a least-squares approach. The task is to calculate the corresponding weight Λ_{dir} from Eqs. (4) and (7) using the microphone signals. To accomplish this, we calculate the pseudo-inverse of a matrix formed by evaluating Eq. (6) for all microphone positions and for all the directions in the measuring grid. This results in the following expression:

$$W_{\text{LS}}(k, \theta_{\text{mic}}, z_{\text{mic}}, \theta_{\text{dir}}, \varphi_{\text{dir}}) = \left[\frac{i \exp[ik \sin\varphi_{\text{dir}} z_{\text{mic}}]}{\pi^2 k \sin\varphi_{\text{dir}} R_{\text{cyl}}} \sum_{n=-M}^{M} \frac{i^n \exp[in(\theta_{\text{mic}} - \theta_{\text{dir}})]}{H_n^{(1)'}(k \sin\varphi_{\text{dir}} R_{\text{cyl}})} \right]^{+}. \tag{10}$$

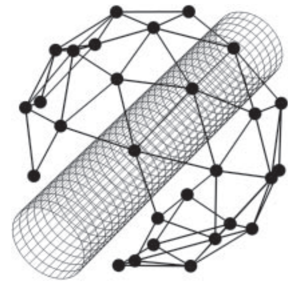

Fig. 4 An illustration of the measuring grid used in our proposal. Each of the points on the imaginary sphere surrounding the cylindrical microphone array represents a direction at which beamforming is carried out to build a spherical pressure distribution. Some points are omitted for clarity, but the measuring grid covers a full sphere around the cylinder.

The microphone positions in cylindrical coordinates are given by the polar angles θ_{mic} and the axial coordinates z_{mic}. The radial component for all microphones is the constant R_{cyl}. Equation (10) defines a spatial filter, W_{LS}, that mixes all of the microphone signals and picks out the sound that arrives from a specific direction of azimuth θ_{dir} and elevation φ_{dir}. This is known as a beamformer. Our proposal uses beamforming to transform the sound pressure distribution on the cylinder, observed by the microphones, into a distribution on the sphere. The spherical distribution is sampled at the directions where beamforming was applied, that is, at all directions represented in the measuring grid. To calculate the values of the spherical distribution, p_{dir}, we apply Eq. (10) to the recorded signals ψ_{t}:

$$p_{\text{dir}}(k) = \sum_{\text{mic}} W_{\text{LS}}(k, \theta_{\text{mic}}, z_{\text{mic}}, \theta_{\text{dir}}, \varphi_{\text{dir}}) \psi_{\text{t}}(k, \theta_{\text{mic}}, z_{\text{mic}}). \tag{11}$$

4.2. Directional encoding and normalization

The spatial encoding stage of our proposal specifies a directional encoding matrix, $\boldsymbol{C}$, defined by its elements as follows:

$$c_{(n+1)\cdot(n+1)-n+m,\mathrm{dir}} = Y_{nm}(\theta_{\mathrm{dir}}, \varphi_{\mathrm{dir}}) \quad \text{for } n < N,$$

$$\left.\begin{aligned} c_{M\cdot M+2n,\mathrm{dir}} &= Y_{n,-n}(\theta_{\mathrm{dir}}, \varphi_{\mathrm{dir}}) \\ c_{M\cdot M+2n+1,\mathrm{dir}} &= Y_{nn}(\theta_{\mathrm{dir}}, \varphi_{\mathrm{dir}}) \end{aligned}\right\} \quad \text{for } M > n > N. \qquad (12)$$

The spherical harmonics are evaluated at all directions represented in the measuring grid. The encoding matrix $\boldsymbol{C}$ has a total of $N^2 + 2M + 1$ rows and as many columns as there are directions in the measuring grid. An encoding of the sound field can be easily calculated from $\boldsymbol{C}$:

$$\tilde{\boldsymbol{B}}(k) = \boldsymbol{C}\boldsymbol{p}(k). \qquad (13)$$

The components of vector $\boldsymbol{p}$ are the results of Eq. (11).

Finally, the proposed encoder normalizes the signals $\tilde{B}$. This step prevents artifacts such as clipping or perceptible quantization noise. Normalization is done by multiplying the corresponding Furse-Malham coefficient [10] N_{nm}^{FM} by each of the signals in $\tilde{B}$. The final encoding produced by our proposal is given by the following expression:

$$B_{nm}(k) = N_{nm}^{\mathrm{FM}} \tilde{B}_{nm}(k). \qquad (14)$$

This normalized coefficients B_{nm} define a MOA encoding of the sound field recorded by the cylindrical microphone array using a full-3D Ambisonic order N and a horizontal Ambisonic order M.

5. Conclusions

In this paper, a new method to record sound fields and encode them using mixed-order Ambisonics was presented. The proposal uses a cylindrical microphone array to sample the sound field. The cylindrical geometry is more natural to sample the horizontal plane to high detail while retaining a low spatial resolution for elevation. The results of the proposed method are conventional MOA encodings that can be reproduced using any loudspeaker array with a suitable decoder.

Acknowledgements

Parts of this research were supported by a Grant-in-Aid for Scientific Research (A) (no. 24240016) to YS.

References

[1] J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, Mass., 1996).

[2] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, **53**, 1004–1025 (2005).

[3] C. Travis, "A new mixed-order scheme for Ambisonic signals," *Proc. Ambisonics Symp.*, 6 pp. (2009).

[4] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, **13**, 135–143 (2005).

[5] S. Koyama, K. Furuya, Y. Hiwasaki, Y. Haneda and Y. Suzuki, "Sound field reproduction using multiple linear arrays based on wave field reconstruction filtering in helical wave spectrum domain," *Proc. ICASSP 2013*, 2842, pp. 271–275 (2013).

[6] R. Duraiswami and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Trans. Audio Speech Lang. Process.*, **18**, 2–16 (2010).

[7] J. Ahonen, G. del Galdo, F. Kuech and V. Pulkki, "Directional analysis with microphone array mounted on rigid cylinder for directional audio coding," *J. Audio Eng. Soc.*, **60**, 311–324 (2012).

[8] S. Holmes, "Circular harmonics beamforming with spheroidal baffles," *Proc. Int. Congr. Acoustics*, 19(055077), 9 pp. (2013).

[9] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, Waltham, Mass., 1999).

[10] D. Malham, *Higher Order Ambisonic Systems. Space in Music — Music in Space*, M.Phil. dissertation, University of York (2003).

[11] J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "Ambisonic synthesis of directional sources using non-spherical loudspeaker arrays," *Proc. 4th Int. Symp. Ambisonics and Spherical Acoust.*, 5 pp. (2012).

[12] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA J. Numer. Anal.*, **19**, 317–334 (1999).