ICA 2016 BUENOS AIRES
Buenos Aires – 5 to 9 September 2016
Acoustics for the 21st Century...

PROCEEDINGS of the 22nd International Congress on Acoustics

Acoustic Array Systems: Paper ICA2016-415

Sound source separation in complex environments using an array-of-arrays microphone system

Jorge Trevino[(a)], Cesar Salvador[(a)], Virgilijus Braciulis[(a)], Shuichi Sakamoto[(a)], Yôiti Suzuki[(a)], Kyoji Yoshikawa[(b)], Takashi Yamasaki[(b)] and Kenichi Kidokoro[(b)]

[(a)]Tohoku University, Japan, jorge@ais.riec.tohoku.ac.jp, salvador@ais.riec.tohoku.ac.jp, virgis@ais.riec.tohoku.ac.jp, saka@ais.riec.tohoku.ac.jp, yoh@riec.tohoku.ac.jp
[(b)]RION Co., Ltd., Japan, yosikawa@rion.co.jp, yamasaki@rion.co.jp, kidokoro@rion.co.jp

Abstract

Recording a specific sound source in a complex environment is challenging, especially when interfering sources lie between the target and the recording system. This can be avoided by placing a microphone next to the target so as to get a clear measurement. However, this may interfere with the events being recorded and lacks flexibility to select and track the desired source. Microphone arrays are used as adaptable systems to separate specific sounds given the spatial position of their sources. Arrays can be classified according to their geometry, such as linear or spherical arrays. In particular, spherical arrays are advantageous due to their compact shape. The highly symmetric design can also yield an almost constant angular resolution making their performance predictable. Unfortunately, this symmetry is also the source of their main limitation: sounds are measured from a single viewpoint. Spherical arrays can separate sounds according to their direction of arrival using a technique known as beamforming; however, it is difficult to separate two sounds when their sources and the array are aligned. The present research introduces a new approach to source separation using an array-of-arrays microphone system. The proposal, a cooperative variant of beamforming, relies on two or more spherical microphone arrays working together as a single, unified system. This yields a multiple-viewpoint measurement of the sound field and allows for sound source separation even when interfering sources are present between the target and each of the arrays. Such separation would be difficult using conventional beamforming. The proposal is evaluated through numerical experiments using both, physical models and real-world measurements for a system composed of two 64-channel microphone arrays.

Keywords: microphone array, sound source separation, beamforming, array-of-arrays

Sound source separation in complex environments using an array-of-arrays microphone system

1 Introduction

Recording of complex environments results in acoustic signals where multiple sound sources are superposed. However, in many situations it is necessary to isolate a specific sound; for example, the speech of a target speaker, or the acoustic cues that can suggest mechanical failures in complex machinery. Microphone arrays are a versatile way to approach this task [1]; their recorded signals can be processed digitally to emphasize sounds originating at a given position while attenuating the rest.

The optimal arrangements of microphones in the array depends on the target application. Several geometries have been considered in previous studies [2, 3]. As an example, spherical microphone arrays are regularly used to record spatial sound; one reason for this is their resemblance to a human head [4, 5, 6]. In addition, the symmetry of spherical acoustic baffles makes it possible to derive an analytic formula for sound scattering around them [7]. Exact models to calculate the expected recorded signals for a spherical microphone array in the presence of a point source are available [4]. Spherical arrays with high channel counts are commercially available, while research prototypes can sometimes use hundreds of microphones [5, 6].

The geometry of spherical microphone arrays allows the use of powerful harmonic analysis techniques [8], leading to spatial sound source separation methods collectively referred to as modal beamforming [9]. These methods can isolate sounds arriving from a specific direction of incidence. This means that the microphone array can be used as if it was a single microphone with a desired directivity. Spherical geometry further allows the directivity pattern to be oriented towards any specified direction.

Conventional beamforming methods cannot isolate a sound source when, from the viewpoint of the microphone array, one or more interfering sound sources are present in the same direction as the target. The present paper outlines a previously proposed approach to the problem of isolating a single sound source in a complex environment [10]. The proposal uses multiple spherical microphone arrays working in synchrony as a unified system to isolate sounds radiating from inside a compact region, as opposed to the extended beams of conventional methods. It is, thus, capable of separating a target sound source even in the presence of interfering sources aligned with it.

The present research evaluates the feasibility and real-world performance of the proposed array-of-arrays sound source separation method. An actual recording system consisting consists of two 64-channel microphone arrays was built and used to record a complex acoustic environment with three sound sources: a target, and two interferers positioned between the target and each of the spherical arrays. The actual recordings are then used in a numerical experiment to determine the spatial selectivity of the proposed method. We find that the proposal is capable of isolating sounds arriving from a target position while limiting the interference from sources present between the target and the recording equipment.

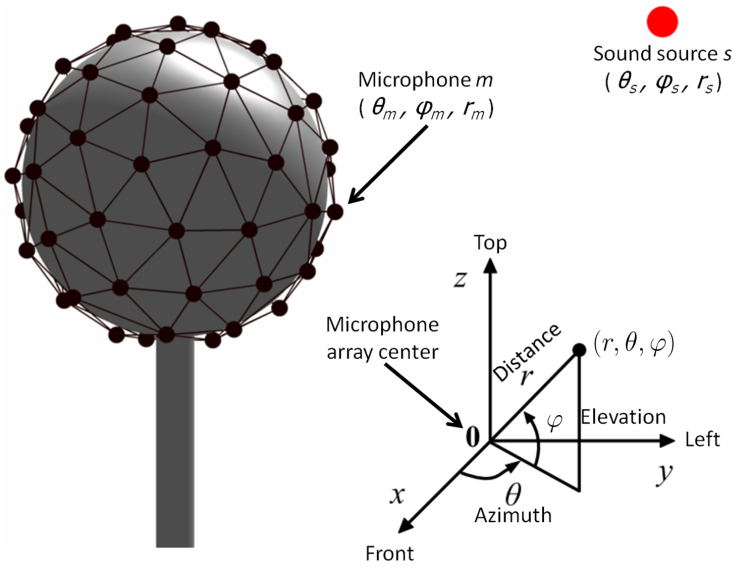

Figure 1: **The spherical coordinate system. This system is typically used to model spherical microphone arrays**

2 Sound source separation using spherical microphone arrays

The symmetry of spherical microphone arrays [7] makes it easier work using a spherical coordinate system like the one shown in Fig. 1. The microphone array is assumed to be positioned at the origin; two angles (azimuth θ and elevation φ) are used to define a direction, and together with a distance r specify a single spatial point.

The body of a spherical microphone array can act as an acoustic baffle, thus affecting the sound pressure measurements and emphasizing their dependency on the direction of arrival for the recorded sounds. The sound pressure observed by a microphone on a rigid sphere in the presence of a point source is given by the following equation [4, 8]:

$$p_{\mathrm{mic}}(k,\vec{r_{\mathrm{src}}}) = p_{\mathrm{src}}(k)\sum_{n=0}^{\infty}(2n+1)\frac{h_n(kr_{\mathrm{src}})}{h_n'(kr_{\mathrm{mic}})}\mathrm{P}_n\left(\frac{\vec{r_{\mathrm{mic}}}\cdot\vec{r_{\mathrm{src}}}}{|\vec{r_{\mathrm{mic}}}||\vec{r_{\mathrm{src}}}|}\right). \quad (1)$$

In this equation, p_{mic} stands for the sound pressure observed at a point $\vec{r_{\mathrm{mic}}}$ on the rigid sphere. A single point source of amplitude p_{src} is assumed to be located at $\vec{r_{\mathrm{src}}}$. Functions h_n and P_n are, respectively, the spherical Hankel functions and the Legendre polynomials, both of order n. Primed symbols are used to denote function derivatives.

2.1 Beamforming using spherical microphone arrays

Beamforming can be used to separate sounds according to their direction of incidence [1, 2, 3, 7, 8, 9]. Separation along distance is generally not attempted and, therefore, beamforming can consider a simplified form of Eq. (1) by removing the terms related to the sound source distance. Alternatively, a reference distance r_{ref} can be used as long as this is large compared

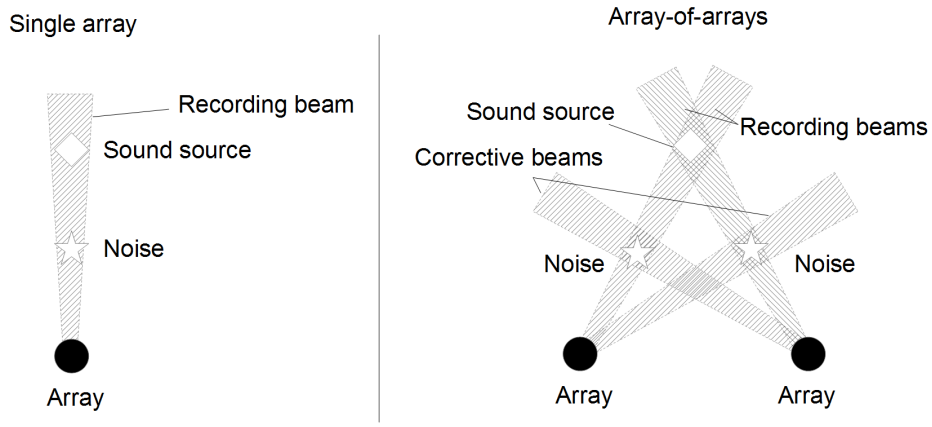

Figure 2: **Illustration of conventional beamforming using a compact microphone array (left) and the proposed sound source separation method using an array-of-arrays (right)**

to the sound source wavelength.

A simple approach to beamforming is to consider a desired directivity pattern $D(\theta,\varphi)$. The target directivity is sampled at a representative set of angles and the sound pressures expected to be observed at each of the microphones due to a sound source at each of the sampled directions is calculated. The recordings for all microphones can then be combined into a single signal approximating the target directivity using the following equation:

$$p_{\text{beam}} = \begin{bmatrix} D_1 & D_2 & \cdots & D_a \end{bmatrix} \begin{bmatrix} p_{1\to1} & p_{2\to1} & \cdots & p_{a\to1} \\ p_{1\to2} & p_{2\to2} & \cdots & p_{a\to2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1\to b} & p_{2\to b} & \cdots & p_{a\to b} \end{bmatrix}^{+} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_b \end{bmatrix} \tag{2}$$

In this equation, a total of a sample directions $D_1,...,D_a$ are sampled using an array of b microphones. The actual microphone recordings are represented by $p_1,...,p_b$, while the ideal sound pressure expected at a given microphone i in the presence of a single sound source at direction j is written as $p_{j\to i}$. This is calculated using Eq. (1) or its plane-wave equivalent. The symbol $|\cdot|^{+}$ denotes a matrix inversion, typically a pseudoinverse.

2.2 Sound source separation using an array-of-arrays

Conventional beamforming using a single microphone array can isolate sounds that originate from a given direction. It, however, cannot separate multiple sources when they are aligned. The problem is hard to overcome since compact arrays can only sample the sound field from a single viewpoint. It is possible, however, to use multiple compact arrays, that is, an array-of-arrays to isolate sounds originating from a specific position in space [11, 12, 13].

Array-of-arrays processing typically considers each of the arrays as an independent beam-

former; they are used to emulate an array of directional microphones. The resulting signals are further processed to isolate the target either through statistical methods [14], or by applying a second beamforming stage designed for complex arrays [15].

Previously, we proposed a new approach to the problem which considers the array-of-arrays system as a unified system [10]. The proposed method is illustrated and compared with conventional beamforming in Fig. 2. The main difference between the proposed method and other existing approaches is that it attempts not only to isolate the target, but also takes advantage of the multiple viewpoints available in the recording system to separate any interfering sound sources and cancel them out in the final output.

Similar to conventional beamforming, the proposal starts by defining a spatial window $W(r,\theta,\varphi)$ which corresponds to the desired selectivity pattern. This is analogous to the target directivity $D(\theta,\varphi)$ used in formulating Eq. (2), however it also includes the radial coordinate. The spatial window is sampled using an a-point grid, with each point corresponding to a possible sound source position $\vec{r}_1,...,\vec{r}_a$. The grid should cover the entire region where sound sources may be present, including the target source.

Equation (1) is once again used to calculate the expected signals at each microphone in each array when a single point source is present at each of the positions sampled by the grid. However, it is important to consider that Eq. (1) assumes the coordinate system of Fig. 1. Since the proposal considers all arrays as a unified system, a single coordinate system with a global origin is required. This leads to the following modification of Eq. (1):

$$H_{\text{mic}}^{\text{array}}(\vec{r_{\text{src}}},k)=\sum_{n=0}^{\infty}(2n+1)\frac{h_n\left[k|\vec{r_{\text{src}}}-\vec{r_{\text{array}}}|\right]}{h_n'(kR_{\text{array}})}\text{P}_n\left(\frac{|\vec{r_{\text{src}}}-\vec{r_{\text{array}}}|^2+|\vec{r_{\text{mic}}}-\vec{r_{\text{array}}}|^2-|\vec{r_{\text{src}}}-\vec{r_{\text{mic}}}|^2}{2|\vec{r_{\text{src}}}-\vec{r_{\text{array}}}||\vec{r_{\text{mic}}}-\vec{r_{\text{array}}}|}\right). \quad (3)$$

The position of the array as seen from the global origin of coordinates is denoted by $\vec{r_{\text{array}}}$, while R_{array} stands for its radius. Vectors $\vec{r_{\text{src}}}$ and $\vec{r_{\text{mic}}}$ retain their original definitions, being defined in the local coordinate systems for each of the arrays.

The proposal can then be summarized by the following equation:

$$p_{\text{target}}(k)=\begin{bmatrix}W(\vec{r}_1) & W(\vec{r}_2) & \cdots & W(\vec{r}_a)\end{bmatrix}\begin{bmatrix}H_1^1(\vec{r}_1,k) & H_1^1(\vec{r}_2,k) & \cdots & H_1^1(\vec{r}_a,k)\\ H_2^1(\vec{r}_1,k) & H_2^1(\vec{r}_2,k) & \cdots & H_2^1(\vec{r}_a,k)\\ \vdots & \vdots & \ddots & \vdots\\ H_b^1(\vec{r}_1,k) & H_b^1(\vec{r}_2,k) & \cdots & H_b^1(\vec{r}_a,k)\\ H_1^2(\vec{r}_1,k) & H_1^2(\vec{r}_2,k) & \cdots & H_1^2(\vec{r}_a,k)\\ H_2^2(\vec{r}_1,k) & H_2^2(\vec{r}_2,k) & \cdots & H_2^2(\vec{r}_a,k)\\ \vdots & \vdots & \ddots & \vdots\\ H_b^N(\vec{r}_1,k) & H_b^N(\vec{r}_2,k) & \cdots & H_b^N(\vec{r}_a,k)\end{bmatrix}^{+}\begin{bmatrix}p_1^1(k)\\ p_2^1(k)\\ \vdots\\ p_b^1(k)\\ p_1^2(k)\\ p_2^2(k)\\ \vdots\\ p_b^N(k)\end{bmatrix} \quad (4)$$

Equation (4) assumes a total of N arrays, each composed of b microphones. The recorded signal for the microphone j in array i is denoted by $p_j^i(k)$. All of these signals are, therefore, combined into a single output which corresponds to a recording with a spatial selectivity approximating the target window function.

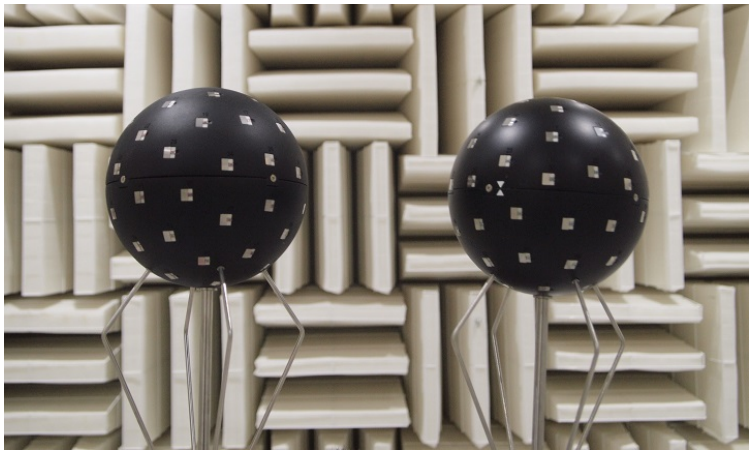

Figure 3: **Two 64-channel spherical microphone arrays used to evaluate the proposal**

3 Evaluation

The proposed array-of-arrays sound source separation method was implemented using two identical spherical microphone arrays. Each of them consists of 64 microphones distributed on the surface of a sphere. The arrays have a radius of 8.5 cm and their distributions were decided according to a relaxation problem for electric charges on the sphere [16]. A photograph of these arrays is shown in Fig. 3.

A major advantage of the almost-regular angular sampling and the spherical geometry is that the performance of the arrays should not be significantly affected by their orientation. Furthermore, since two arrays are used to isolate a target sound source, the entire recording environment can be reduced to the plane defined by these three positions. Any interfering sound source lying outside this plane will not be aligned with the target and the microphones and, therefore, can be easily discriminated through conventional beamforming methods. For these reasons, evaluation conditions will be limited to a 2-dimensional plane.

To consider the worst-case situation, two interfering sources were added, each between one of the microphone arrays and the target. The configuration of the recording environment is shown in Fig. 4. Parameters α and d were studied through computer simulations. The performance of the proposal was found to be stable for all angles between approximatly 40° and 150°. Similarly, no significant changes in the performance were observed once the edge length exceeded 1 meter. For our measurements, we decided to set the angle α to 60°, thus resulting in an equilateral triangle. The length of the triangle's edges d was set to 2 meters. The grid used to define a target spatial selectivity was limited to a 2-by-2 meters square, with positions sampled regularly every 10 cm. All sound sources were located inside the grid, with the microphone arrays located 70 cm away from their closest grid points. Once again, these parameters were selected through numerical experiments to show the typical performance of the system.

Figure 5 shows the experimental setup in an anechoic chamber. All sources radiated uncorrelated white noise at the same sound pressure level. Therefore, recordings with a single

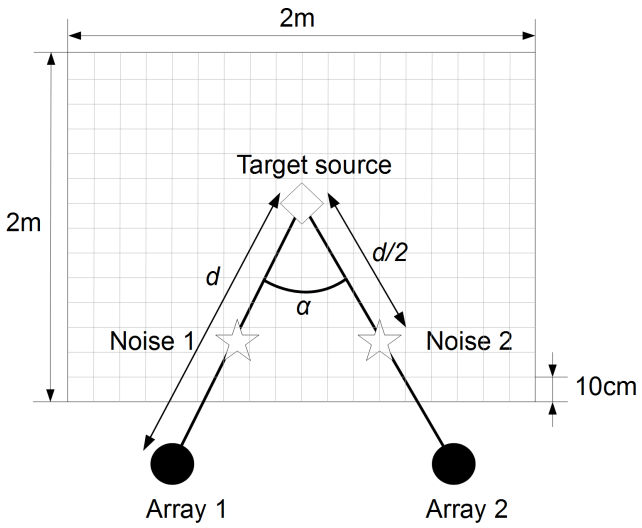

Figure 4: **The recording environment. Two microphone arrays and one target sound source as positioned at the corners of an isosceles triangle. Two interferers lie between each of the arrays and the target, and therefore on the plane of the triangle**

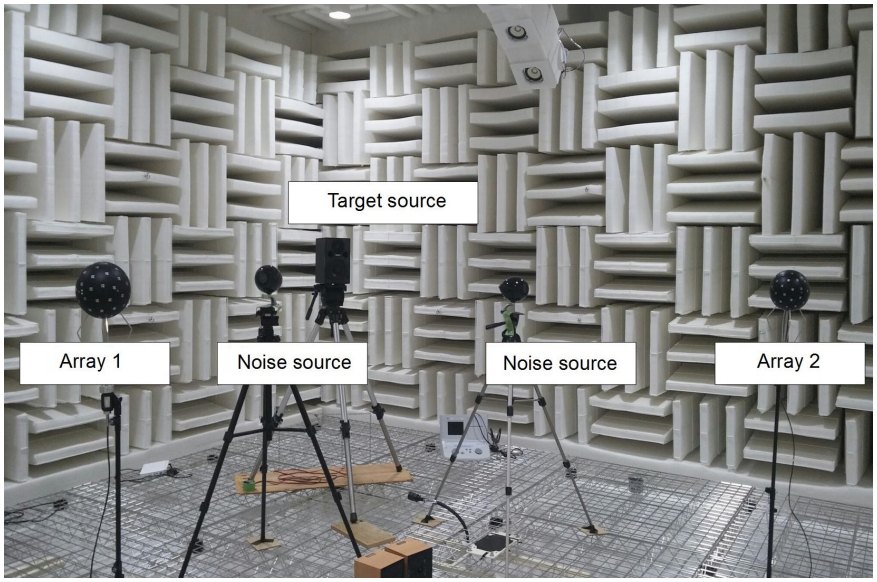

Figure 5: **The recording environment used in our evaluation. Two 64-channel spherical microphone arrays and a target sound source are located at the vertices of an equilateral triangle with an edge length of 2 meters. Interfering sound sources are placed half-way between each microphone array and the target. The whole system was assembled inside an anechoic chamber**

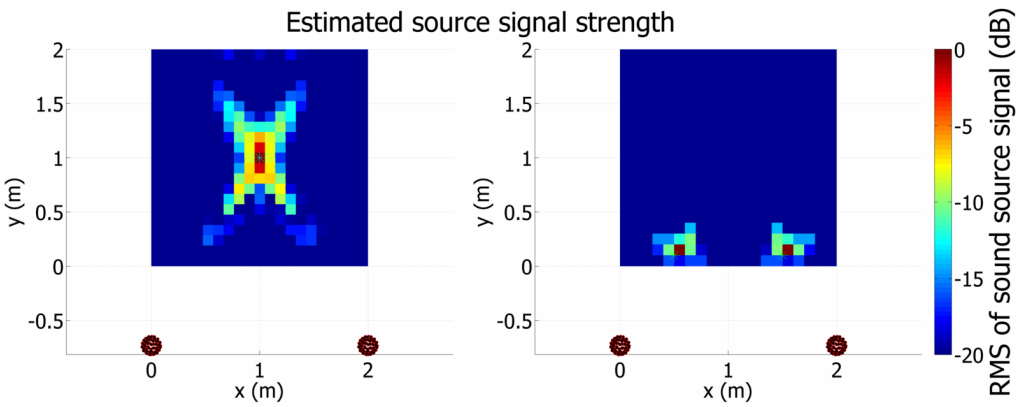

Figure 6: **Distribution of acoustic energy on the measurement grid, as identified by the proposed method. The spatial signature of the target is shown on the left panel, with most of its energy concentrated along a cross centered at its actual position. The right panel shows the spatial signature of the interfering sources, showing negligible leakage at the target position**

microphone placed at one of the array positions would exhibit a signal-to-noise ratio (SNR) of -12 dB. Recordings for the interferers and the target were done independently to identify their spatial signature on the grid.

The proposed algorithm was repeatedly applied to the microphone recordings using a delta function as target spatial window. The delta was positioned at each of the grid points and the energy (root-mean-square) for each output was calculated. This data was put together to construct a spatial signature for the sound sources, as inferred by the proposed method. Figure 6 shows the spatial signatures obtained for the target sound source, as well as the interfering sources.

Due to linearity, applying the proposal to recordings done in the presence of all sound sources would result in the sum of the two distributions shown in Fig. 6. The distinct spatial separation of all sound sources means that any spatial window covering the deeper part of the grid (beyond $y = 0.5$ m while discarding the rest will be able to separate the target. Simulation results for this particular case, show the expected SNR to exceed 35 dB, an improvement of almost 50 dB from the SNR present at the input.

4 Conclusions

A method to record specific sound sources in a complex environment using an array-of-arrays system was reviewed and evaluated. An actual recording system was built using two 64-channel spherical microphone arrays. The system was used to record target and interfering sound sources in an anechoic chamber, and the proposed sound source separation method was applied to these recordings repeatedly. Spatial signatures for all sound sources were in-

ferred from the array-of-arrays system. These show a distinct separation between the target and the interferers and, therefore, imply that sound source separation is possible by applying a simple spatial window. These experimental results confirm those of numerical simulations which show that the proposed method can effectively improve the signal-to-noise ratio of the input signals by almost 50 dB in ideal conditions. Future work will include the evaluation of the SNR in experimental conditions which include noise and microphone positioning errors.

Acknowledgements This work was partly supported by a JSPS KAKENHI Grant-in-Aid for Scientific Research (A) (No. 16H01736) to S.Y. and the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet."

References

[1] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustics Wavefield Decomposition,* Springer, 2007.

[2] D.N. Zotkin, R. Duraiswami and N.A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Trans. on Audio, Speech and Lang. Proc.*, Vol. **18** (1), pp. 2–16, 2010.

[3] S. Holmes, "Circular harmonics beamforming with spheroidal baffles," *Proc. Int. Congr. on Acoust.*, POMA **19** (055077), pp. 1–9, 2013.

[4] R.O. Duda and W.L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, Vol. **104** (5), pp. 3048–3058, 1998.

[5] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya and Y. Suzuki, "Sound-space recording and binaural presentation system based on a 252ch microphone array," *Acoust. Sci. and Tech.*, Vol. **36** (6), pp. 516–526, 2015.

[6] 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," *Proc. 2nd Int. Symp. on Univ. Comm.*, pp. 429–434, 2008.

[7] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. on Speech and Audio Proc.*, Vol. **13** (1), pp. 135–143, 2005.

[8] E.G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography,* Academic Press, 1999.

[9] A.M. Torres, M. Cobos, B. Pueo and J. Lopez, "Robust acoustic source localization based on modal beamforming and time-frequency processing using circular microphone arrays," *J. Acoust. Soc. Am.*, Vol. **132** (3), pp. 1511–1520, 2012.

[10] J. Trevino, S. Sakamoto and Y. Suzuki, "Separation of spatially-segregated sound sources using multiple spherical microphone arrays,"*Proc. Spring Meeting of the Acoustical Society of Japan* 1-Q-44 (in CD-ROM), 2015.

[11] G. del Galdo, O. Thiergart, T. Weller and E.A.P. Habets, "Generating virtual microphone signals using geometrical information gathered by distributed arrays," *Joint Workshop on Hands-free Speech Comm. and Mic. Arrays*, pp. 185–190, 2011.

[12] F. Wang and X. Pan, "A Novel Algorithm for Wideband Acoustic Sources Localization Using Multiple Spherical Arrays," *Proc. Int. Symp. Signal Proc. and Inf. Tech.*, pp. 249–254, 2013.

[13] M. Compagnoni, P. Bestagini, F. Antonacci, A. Sarti and S. Tubaro, "Localization of Acoustic Sources Through the Fitting of Propagation Cones Using Multiple Independent Arrays," *IEEE Trans. on Audio, Speech and Lang. Proc.*, Vol. **20** (7), pp. 1964–1975, 2012.

[14] P. Comon, "Independent component analysis, A new concept?," *Sig. Proc.*, Vol. **36**, pp. 287–314, 1994. P. Comon, *Sig. Proc.*, **36**, pp. 287–314, 1994.

[15] I. Tashev and H.S. Malvar, "A new beamformer design algorithm for microphone arrays," *30th Proc. IEEE Int. Conf. on Acoust. Speech and Sig. Proc.*, pp. 101–104, 2005.

[16] V.A. Yudin, "The minimum of potential energy of a system of point charges," *Discrete Math. Appl.*, Vol. **3** (1) pp. 75–81, 1993.