# PHYSICAL-MODEL BASED EFFICIENT DATA REPRESENTATION FOR MANY-CHANNEL MICROPHONE ARRAY

*Yuji Koyano, Kohei Yatabe, Yusuke Ikeda and Yasuhiro Oikawa*

Department of Intermedia Art and Science, Waseda University, Tokyo, Japan

## ABSTRACT

Recent development of microphone arrays which consist of more than several tens or hundreds microphones enables acquisition of rich spatial information of sound. Although such information possibly improve performance of any array signal processing technique, the amount of data will increase as the number of microphones increases; for instance, a 1024 ch MEMS microphone array, as in Fig. 1, generates data more than 10 GB per minute. In this paper, a method constructing an orthogonal basis for efficient representation of sound data obtained by the microphone array is proposed. The proposed method can obtain a basis for arrays with any configuration including rectangle, spherical, and random microphone array. It can also be utilized for designing a microphone array because it offers a quantitative measure for comparing several array configurations.

***Index Terms***— MEMS microphone, principal component analysis (PCA), wave equation, solution space, plane wave.

## 1. INTRODUCTION

In contrast to a microphone which acquires information only at a single point, a microphone array is a fundamental device for observing spatial information of a sound field at multiple points. There are a lot of signal processing methods effectively utilizing spatial information of sound for several applications including direction-of-arrival estimation [1, 2], noise reduction [3, 4], and blind source separation [5–7]. While the most widely used microphone array is a two-channel one because it is ubiquitous, an array consists of more than two microphones is getting more and more popular today [8–10].

Recently, development of a large-scale recording system, possibly combined with micro electro mechanical systems (MEMS) technology, allows construction of a microphone array with more than several tens or hundreds microphones [11–19]. Such many-channel microphone array has possibility of improving performance of most of array processing technique by providing rich spatial information. However, large number of microphones produce huge amount of data that may not be tractable for practical situations. For example, a 1024 ch MEMS microphone array, produced by Takeoka, Oikawa *et al.* in 2010, shown in Fig. 1 generates more than 10 GB par minute; it is not easy to store and manage that much data. Therefore, some efficient representation of the huge data is desired because spatial information of sound is essentially limited by its (time-directional) frequency.

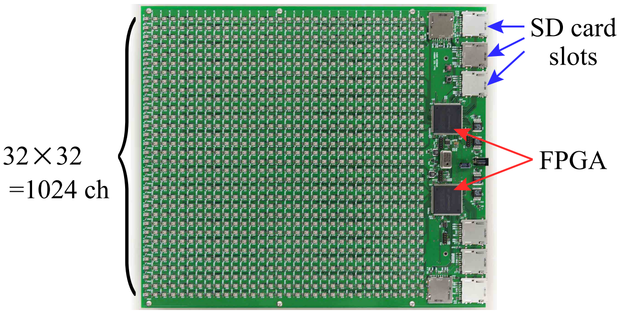


**Fig. 1**. A 1024 ch array board on which MEMS microphones are placed with 1 cm interval. Each microphone outputs $\Delta\Sigma$-modulated binary signal which is directly processed in the stand-alone FPGAs and stored in 16 SD cards simultaneously.

One such popular representation is based on the Fourier basis including the plane wave decomposition, for a one- or two-dimensional array, and the spherical harmonic expansion, for a spherical microphone array; plane waves are the Fourier basis for the Cartesian coordinate system, and spherical harmonics are the Fourier basis for the (angular portion of) Spherical coordinate system. Storing the expansion coefficients of the above orthogonal basis has two advantages: (1) orthogonality ensures the non-redundant representation; and (2) it is easy to analyze the approximation error caused by truncation of the coefficients for data compression. However, these predetermined bases can only be applied to microphone arrays constructed on the very specific coordinate systems; it is not easy to derive an orthogonal basis analytically for, say, a randomly placed microphone array, which should offers better information than regularly placed microphones as well known from the research of compressed sensing.

In this paper, a method for constructing the efficient orthogonal basis for representing sound information obtained by a microphone array whose geometry can be arbitrary is proposed. The proposed method applied principal component analysis (PCA) to artificial data generated so that they span only a subset of the solution space of the wave equation. Since the basis is constructed via the physical model, the strong correlation of sound information in time and space is efficiently took into account. In addition, as a by-product of representing sound information by a model-based orthogonal basis, the method can also be used for comparing several array configurations quantitatively; it is not easy to compare, for instance, rectangle, spherical and random microphone arrays all at once because the spatial sampling theory generally depends on the configuration of sampling points.

## 2. SOLUTION SPACE OF WAVE EQUATION AND ITS PLANE WAVE REPRESENTATION

In linear acoustics, sound propagation is described by the wave equation,

$$\left( \triangle - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) p(\boldsymbol{x}, t) = 0, \tag{1}$$

where $\triangle = \sum_n \partial^2 / \partial x_n^2$ is the Laplace operator, $t$ is time, $\boldsymbol{x} = (x_1, x_2, x_3)$ is position, $p$ is sound pressure, and $c$ is the speed of sound. The Fourier transform on time variable $\mathscr{F}_t$ converts Eq. (1) into the Helmholtz equation:

$$\left( \triangle + k^2 \right) u(\boldsymbol{x}, \omega) = 0, \tag{2}$$

where $k = \omega / c$ is the wave number, and $\omega$ is the angular frequency. Any solution $u$ to the homogeneous Helmholtz equation can be approximated arbitrarily well by the linear combination of plane waves [20]:

$$u(\boldsymbol{x}, \omega) = \sum_n \alpha_n \exp(jk \langle \boldsymbol{x}, \boldsymbol{\nu}_n \rangle), \tag{3}$$

where $\alpha_n \in \mathbb{C}$, $j = \sqrt{-1}$, $\langle \cdot, \cdot \rangle$ is the standard inner product, $\boldsymbol{\nu} \in \mathbb{S}^2$ is unit vector which corresponds to direction of propagation of the plane wave.

Let us consider the spatial Fourier transform $\mathscr{F}_{\boldsymbol{x}}$ of $u$,

$$\mathscr{F}_{\boldsymbol{x}} : u(\boldsymbol{x}, \omega) \mapsto U(\boldsymbol{k}, \omega), \tag{4}$$

where $\boldsymbol{k} = (k_x, k_y, k_z)$. It is easy to see that a plane wave is mapped to the Dirac delta function by the transform:

$$\mathscr{F}_{\boldsymbol{x}} \big[ \exp(jk \langle \boldsymbol{x}, \boldsymbol{\nu}_n \rangle) \big] = \delta(\boldsymbol{k}_n, \omega). \tag{5}$$

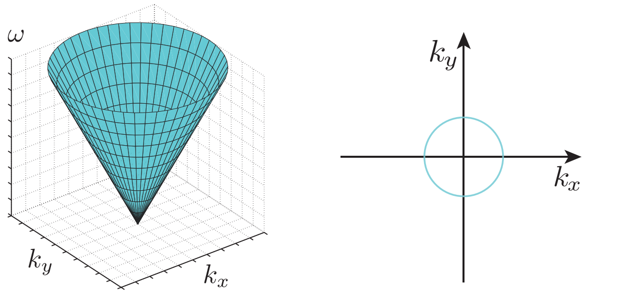
Therefore, the spatio-temporal Fourier transformed solution $U$ can be represented as

$$U(\boldsymbol{k}, \omega) = \mathscr{F}_{\boldsymbol{x}} \Big[ \sum_n \alpha_n \exp(jk \langle \boldsymbol{x}, \boldsymbol{\nu}_n \rangle) \Big] = \sum_n \alpha_n \delta(\boldsymbol{k}_n, \omega), \tag{6}$$

where

$$\boldsymbol{k}_n \in \left\{ \boldsymbol{k} \in \mathbb{R}^3 \mid k_x^2 + k_y^2 + k_z^2 = k^2 \; (= \omega^2 / c^2) \right\}. \tag{7}$$

Thus, by ignoring the evanescent wave component because it exponentially decays with distance from the sound source, the solution to the wave equation only lies on the corn illustrated in Fig. 2 (this figure is for the two-dimensional wave equation because it is difficult to depict three-dimensional case which end up with four-dimensional spectrum) [21].



(a) Region where $U_{\mathrm{2D}}$ lies. (b) Cross-sectional view of (a).

**Fig. 2**. Schematic diagram of the region where the solution to the two-dimensional homogeneous Helmholtz solution exist in the spatio-temporal frequency domain (evanescent component of the solution is neglected in this paper).

**Table 1**. Machine epsilon for several data format. The $\varepsilon_p$ for integer type are defined here based on each maximum representable number $N_{\max}$ as $\varepsilon_{\text{int}} = 1/N_{\max}$.

| Numeric data format | Machine epsilon $\varepsilon_p$ |
|---|---|
| 32bit-float | $1.192 \times 10^{-7}$ |
| 64bit-float | $2.220 \times 10^{-16}$ |
| 16bit-integer | $3.051 \times 10^{-5}$ |
| 24bit-integer | $1.192 \times 10^{-7}$ |
| 32bit-integer | $4.656 \times 10^{-10}$ |

## 3. EFFICIENT REPRESENTATION OF MICROPHONE ARRAY SIGNALS

In the previous section, we see that a sound field, or a solution to the wave equation, has highly localized spatio-temporal spectrum. Therefore, a spanning set of such spectrum can effectively represent the information related to sound contained in observation signals of a microphone array. However, construction of the spanning set for a microphone array signal which depend on the array configuration is not trivial. For instance, spatio-temporal spectrum of a flat array as in Fig. 1 can be characterized by the convolution of the cone and the sinc function corresponding to the spatial rectangular window, but how about an array consists of randomly placed microphones? In this paper, a general framework for constructing the efficient basis regardless of array configuration is proposed based on the physical model explained in Sec. 2.

### 3.1. Proposed method

A sound field $u(\boldsymbol{x}, \omega)$ is discretized spatially by an $M$ channel microphone array as $\{u(\boldsymbol{x}_m, \omega)\}_{m=1}^{M}$. This discretization corresponds to a mapping from infinite dimensional vector space to an $M$ dimensional subspace. For efficient data representation, a basis for the $M$ dimensional space is desired.

**Table 2**. Summary of array configurations used for the experiment, each character corresponds to the same one in Fig. 3, 4 and 6. Size of the arrays are determined so that: (a), (b), (d) and (f) have equal microphone density per surface area; (c), (d), (e) and (h) have equal microphone density per volume; and radius of the spherical arrays (f) and (g) are set to equal. All random arrays are generated from the uniform distribution inside each region. All configurations consist of 1024 ch microphones.

| | |
|---|---|
| (a) | Flat square microphone array with area of 0.0961 m$^2$ ( $=$ 0.31 m $\times$ 0.31 m) defined on a square lattice pattern at even intervals of 0.01 m as in Fig. 1. |
| (b) | Randomly placed microphone array inside a square with area of 0.0961 m$^2$ ( $=$ 0.31 m $\times$ 0.31 m). |
| (c) | 3D array that microphones are placed inside a rectangular region with volume of 0.0961 m$^2$ ( $=$ 0.458 m $\times$ 0.458 m $\times$ 0.458 m) defined on a square lattice pattern at even intervals of 0.01 m. |
| (d) | 3D array that microphones are placed on a surface of a rectangular region with volume of 0.0961 m$^2$ ( $=$ 0.458 m $\times$ 0.458 m $\times$ 0.458 m) defined on a square lattice pattern at even intervals of 0.01 m. |
| (e) | Randomly placed microphone array inside a rectangular region with volume of 0.0961 m$^2$ ( $=$ 0.458 m $\times$ 0.458 m $\times$ 0.458 m). |
| (f) | Spherical array that microphones are placed on a sphere with the radius of 0.0903 m at quasi-equal intervals. |
| (g) | Randomly placed microphone array inside a sphere with the radius of 0.0903 m. |
| (h) | Randomly placed microphone array inside a sphere with the radius of 0.056 m. |

However, characteristic of the space depends on the array configuration; analytical approach for constructing the basis has to be specific for each configuration, such as plane or spherical array. Instead, a numerical method combined with artificial data generation is presented here.

Since a sound field in a region of interest, which does not contain any sound source in it, can be represented by plane waves as in Eq. (3), every sampled data can also be represented as

$$u(\boldsymbol{x}_m, \omega) = \sum_n \alpha_n \exp(jk\langle \boldsymbol{x}_m, \boldsymbol{\nu}_n\rangle). \tag{8}$$

While sampled data $\{u(\boldsymbol{x}_m, \omega)\}$ depends on microphones' position $\{\boldsymbol{x}_m\}$, the plane waves themselves in the right hand side only depend on $\boldsymbol{\nu}_n$. Therefore, any data measured by a microphone array can be reproduced by a superposition of measured plane waves $\{\exp(jk\langle \boldsymbol{x}_m, \boldsymbol{\nu}_n\rangle)\}_{n=1}^{N}$.

Based on the above theoretical reasoning, we propose a basis construction method using PCA on artificially generated microphone array signals from the set of plane waves. The proposed method is summarized as follows:

1. Import microphone array configuration $\{\boldsymbol{x}_m\}_{m=1}^{M}$.
2. Create plane wave data $\boldsymbol{p}_n = \{\exp(jk\langle \boldsymbol{x}_m, \boldsymbol{\nu}_n\rangle)\}_{m=1}^{M}$ from the configuration where $\{\boldsymbol{\nu}_n\}_{n=1}^{N}$ is sampled from the unit sphere $\mathbb{S}^2$.
3. Apply PCA to the generated data $\{\boldsymbol{p}_n\}_{n=1}^{N}$.
4. Adopt principle components corresponding to the eigenvalues larger than predetermined threshold as a basis for representing microphone array signals.

The truncation threshold in step 4 can be reasonably chosen by desired accuracy of numerical approximation. Here, machine epsilons in Table 1 are utilized as the threshold because they decide final accuracy of the representation.

From the construction, proposed basis only spans the space where propagating waves exist. Therefore, it can not only be used for efficient representation but also reduce measurement noise for low frequency sound because orthogonal projection to such space eliminates any component which does not behave as sound. In addition, the eigenvalues provides a common performance measure that how efficiently the microphone array obtains spatial information, regardless of the configuration of a microphone array.

## 4. NUMERICAL EXPERIMENT

It was investigated by an experiment that how efficiently the model-based orthogonal basis obtained by the proposed method can represent data. The experiment also confirmed that the proposed method can also be used as a performance index of microphone array configurations.

For the experiment, we considered 8 microphone arrays with different configurations as in Table 2 and Fig. 3. $\{\boldsymbol{\nu}_n\}$ in Eq. (8) was obtained by sampling 16384 points from the unit sphere quasi-equally for generating plane waves as learning data $\{\boldsymbol{p}_n\}_{n=1}^{16384}$.

Figure 4 shows an example of eigenvalues, normalized by the largest eigenvalue, only considering a 5 kHz sound field with the sound speed 340 m/s as illustrated in Fig. 2(b); the horizontal purple lines indicate the machine epsilons in Table 1 that are used for criteria on basis truncation. Examples of basis elements obtained for the flat array (configuration (a)) are depicted in Fig. 5. Figure 6 shows the percentage of the number of the basis needed to represent data over the number of original data (it is integrated value over all bandwidth up to the value indicated by the horizontal axis, for instance, configuration (a) can acquire spatial information of sound signal,

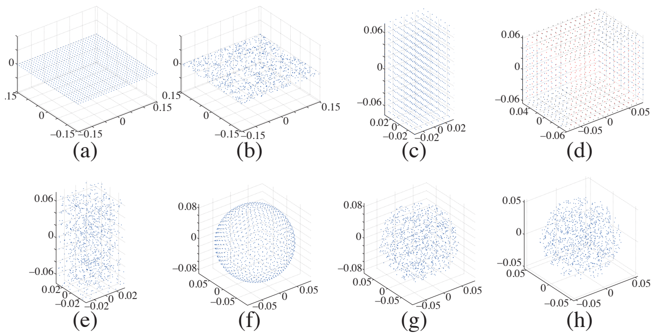


**Fig. 3**. Illustration of the array configurations used in the experiment. Each character corresponds to the one in Table 2.

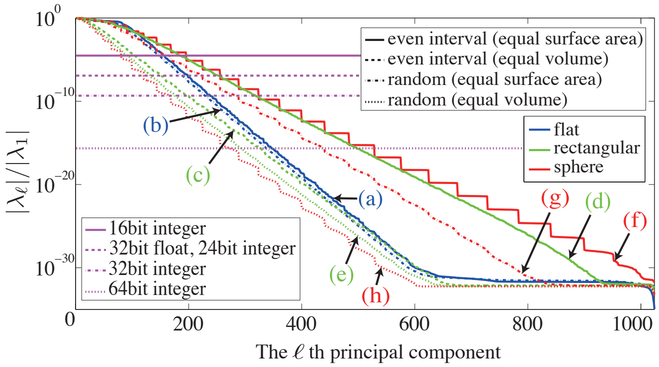


**Fig. 4**. Eigenvalues of principle components normalized by the largest eigenvalue $\lambda_1$. Each color represents the shape of region containing microphones while the type of lines represent the construction rule.

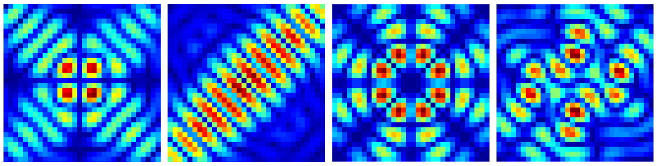


**Fig. 5**. Examples of basis elements obtained from a 5 kHz sound field with the sound speed 340 m/s for the array (a).

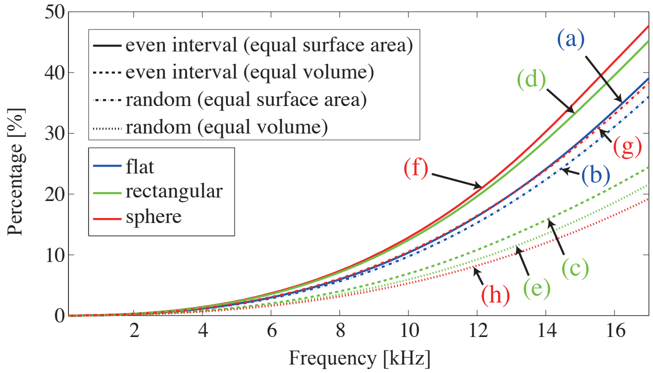


**Fig. 6**. Percentage of the number of the basis needed to represent data with 16 bit accuracy over the original data. The horizontal axis indecates the maximum frequency of a wideband signal that can be represented by the basis.

which contains frequency component of 20 Hz through 10 kHz, by only 10 % of the data size compared to the original size). From Fig. 6, it is confirmed that when we want to record spatial information of sound up to 17 kHz, for any configuration, the proposed basis can represent the signal by less than 50 % data size, and for the best configuration, less than 20 % (as mentioned in the previous section, the redundant information is effective to improve noise robustness).

Interestingly, array configurations with the same microphone interval or the same microphone density yield different results. For example, although microphones for (a), (c) and (d) are all placed on the rectangular coordinate with 1 cm interval, the flat array (a) needs about 1.5 times more basis elements comparing to the 3D rectangular array (c), and the rectangular surface array (d) needs around 2 times more [1]. Moreover, random arrays obtained spatial information more efficiently than regular arrays; this phenomenon agrees with the previous research that a random sampler gives better information than a regular one [22]. The curve in Fig. 6 can be considered as the sampling property of microphones arrays, i.e., the frequency that the curve for an array intersect with the 100 % line can be considered as the spatial Nyquist frequency. In other words, an array configuration with the lower curve obtains spatial information of sound more efficiently. Thus, the proposed method can be used as performance indicator for microphone arrays without knowledge of the sampling theory specific for each configuration.

## 5. CONCLUSION

In this paper, we proposed a general framework for constructing an orthogonal basis for efficiently representing sound information obtained by a microphone array whose geometry can be arbitrary. The proposed method is based on the physical model of sound so that the basis spans a subset of the solution space of the wave equation only. From the results of the numerical experiment, it is confirmed that the proposed basis can reduce data size depending on the array configuration and desired bandwidth. The representation also has an advantage: by using redundant information in oversampled data, measurement noise, which does not behave as sound, is reduced by the orthogonal projection to the solution space. In addition, as demonstrated in Sec. 4, the proposed method can also be used as a performance indicator for designing a microphone array. Since it reveals the sampling property of an array with any configuration, each microphone array can be compared without configuration-dependent theories.

In real-life use of a microphone array, there may be a bias in the arrival direction of the sound source. Therefore, improving the method by including the bias in the arrival direction should be considered in the future.

[1] We noticed later that comparing arrays lying in different dimensional spaces (1D, 2D or 3D) needs additional care. However such difference can be detected easily by perturbating positions of microphones and analyzing the eigenvalues. Thus, it may not be a critical issue in practice.

## 6. REFERENCES

[1] D.B. Ward, Z. Ding, and R. Kennedy, "Broadband DOA estimation using frequency invariant beamforming," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1463–1469, 1998.

[2] K. Kumatani, J. McDonough, and B. Raj, "Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 127–140, 2012.

[3] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, 2002.

[4] R. Miyazaki, H. Saruwatari, S. Nakamura, K. Shikano, K. Kondo, J. Blanchette, and M. Bouchard, "Musical-noise-free blind speech extraction integrating microphone array and iterative spectral subtraction," *Signal Process.*, vol. 102, pp. 226–239, 2014.

[5] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[6] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency binwise clustering and permutation alignment," vol. 19, no. 3, pp. 516–527, 2011.

[7] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," in *Int. Conf. Acoust., Speech Signal Process. (ICASSP)*. IEEE, 2015, pp. 276–280.

[8] Y. Zhang and J. Chambers, "Exploiting all combinations of microphone sensors in overdetermined frequency domain blind separation of speech signals," *Int. J. Adapt. Control Signal Process.*, vol. 25, no. 1, pp. 88–94, 2011.

[9] C. Osterwise and S.L. Grant, "On over-determined frequency domain BSS," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 5, pp. 956–966, 2014.

[10] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Relaxation of rank-1 spatial constraint in overdetermined blind source separation," in *Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 1271–1275.

[11] H.F. Silverman, W.R. Patterson, and J.L. Flanagan, "The huge microphone array," *IEEE Concurr.*, vol. 6, no. 4, pp. 36 – 46, 1998.

[12] S. Enomoto, Y. Ikeda, S. Nakamura, and S. Ise, "Three-dimensional sound field reproduction and recording system based on boundary surface control principle," in *14th Int. Conf. Auditory Disp. (ICAD)*, 2008.

[13] A. Omoto and I. Ikeda, "Construction of 80-channel mobile sound recording system," in *AES Jpn. Sect. Conf. Sendai*, 2012.

[14] A. Omoto, S. Ise, Y. Ikeda, K. Ueno, S. Enomoto, and M. Kobayashi, "Sound field reproduction and sharing system based on the boundary surface control principle," *Acoust. Sci. & Tech.*, vol. 36, no. 1, pp. 1–11, 2015.

[15] T. Okamoto, R. Nishimura, and Y. Iwaya, "Estimation of sound source positions using a surrounding microphone array," *Acoust. Sci. & Tech.*, vol. 28, no. 3, pp. 181–189, 2007.

[16] Y. Suzuki, T. Okamoto, J. Trevino, Z.-L. Cui, Y. Iwaya, S. Sakamoto, and M. Otani, "3D spatial sound systems compatible with human's active listening to realize rich high-level kansei information," *Interdiscip. Inf. Sci.*, vol. 18, no. 2, pp. 71–82, 2012.

[17] 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 *20th Int. Congr. Acoust. (ICA)*, 2010.

[18] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya, and Y. Suzuki, "Sound-space recording and binaural presentation system based on a 252ch microphone array (in press)," *Acoust. Sci. & Tech.*

[19] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 205 – 216, 2014.

[20] A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous helmholtz solutions," *Z. Angew. Math. Phys.*, vol. 62, pp. 809–837, 2011.

[21] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, 2006.

[22] G. Chardon, A. Cohen, and L. Daudet, "Sampling and reconstruction of solutions to the Helmholtz equation," *Sampl. Theory Signal Image Process.*, vol. 13, no. 1, pp. 67–90, 2014.