Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **6**, November **2017**

Enhancement of Spatial Sound Recordings by Adding Virtual Microphones to Spherical Microphone Arrays

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, Miyagi, 980-8577, Japan
salvador@ais.riec.tohoku.ac.jp; saka@ais.riec.tohoku.ac.jp;
jorge@ais.riec.tohoku.ac.jp; yoh@riec.tohoku.ac.jp

Received January 2017; revised May 2017

Abstract. *Spherical microphone arrays mounted on a rigid spherical baffle effectively capture acoustic environments for their reconstruction by sensing the space in all directions. The array signals are further encoded for their scalable processing using the spherical Fourier transform. Recent spatial sound applications are demanding arrays with a large number of microphones. However, physically increasing the spatial resolution of available arrays is not always feasible. In environments such as conference rooms or concert halls the source positions are often confined to a small region of space. When prior knowledge about source positions is assumed, the pressure generated at any point on the baffle can be estimated with a physical model of the rigid sphere. In this paper, the rigid sphere model is used to define a surface pressure variation function that relates the pressure at two arbitrary points on the baffle. Based on this function, a spatial resolution enhancement method for spherical arrays is proposed, which aims to add virtual microphones to the array by synthesizing recording of signals at positions without microphones. The proposal constitutes a preprocessing stage intended to be applied before array signal encoding. Numerical experiments show that the enhancement of spatial resolution is possible all over the sphere if the number of real microphones is sufficiently large.*
Keywords: 3D audio technology, Sound field recording, Sound field interpolation, Spherical microphone arrays, Array signal processing, Spherical acoustics.

1. **Introduction.** The spatial features of sound, such as the reverberation of the environment and the positions of sound sources, play an essential role when presenting three-dimensional (3D) information to users in realistic multimedia systems. Spatial sound indeed enhances the levels of perceived presence and naturalness because it increases the awareness of physical surroundings [1, 2], improves the perception of the semantic and emotional components of daily experience [3, 4], and contributes to the perception of self-motion [5, 6, 7, 8, 9]. Spatial sound technologies for the enrichment of audio content are therefore key components in the development of future multimedia telecommunication systems that aim to re-create the perceptual experience of being immersed in a distinct environment. Such immersive systems are, in turn, important to enable more realistic telepresence applications for remote collaboration between users [10].

Several methods for presenting 3D sound through loudspeaker arrays, personal sound zone devices, and headphones have been proposed over the last decades (see [11, 12, 13, 14] for present accounts on these methods). These methods have been mainly used for the spatial composition of auditory scenes based on the enrichment of monophonic audio

contents by providing them with spatial features. When aiming for the proper reconstruction of acoustic environments, however, immediate support for the direct recording of spatial sound is not always available for existing 3D presentation systems. This can be attributed to the diverse application-oriented criteria used to distribute microphones in the recording environment [15, 16]. The proper selection of microphone array topology is therefore crucial to ensure the accurate reconstruction of acoustics environments for immersive multimedia telecommunications.

Spherical microphone arrays are an effective and flexible tool to capture acoustic environments for their reconstruction. Spherical arrays are especially useful in applications that require the capture of sound from all directions within the vicinity of a single vantage point. This is the case, for example, in 3D sound methods that aim to reconstruct sound pressure signals in the ears of listeners [14, 17, 18, 19, 20, 21, 22, 23]. The use of uniform spherical distributions of microphones further enable flexible array signal processing at scalable angular resolutions in a transform domain provided by the spherical Fourier transform [24, 25]. The use of an acoustically rigid baffle to mount the microphones adds stability to the reconstruction. Moreover, transform-domain representations are the basis for a spatial sound encoding format known as high-order ambisonic (HOA) [26, 27, 28].

The capture of high-definition spatial sound to ensure the accurate reconstruction of an acoustic environment demands spherical arrays with a large number of microphones [14, 17, 18, 19, 20, 21, 22, 23]. The costs involved in the construction of high spatial resolution arrays have confined their use to research purposes though [18, 20, 21], and the maximum spatial resolutions of commercially available arrays are still in the range of a few tens of microphones [24, 25]. Moreover, physically increasing the spatial resolution of available arrays is not always feasible. It would be useful, then, if an approach existed for increasing the spatial resolution of available spherical microphone arrays by means of array signal processing methods. Some conditions that make the formulation and application of such an approach possible are considered below.

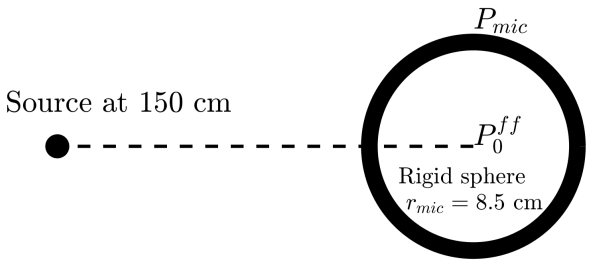
In environments such as conference rooms or concert halls the positions of sound sources are often confined to a small region of space. In these conditions, when the sound source positions can be assumed to be known, the pressure generated at any point on the rigid spherical baffle where the microphones are mounted can be estimated using a physical model of the rigid sphere [29, 30, 31].

In this paper, we propose a spatial resolution enhancement method for rigid spherical microphone arrays in the above-mentioned conditions. The method relies on adding virtual microphones to the array by synthesizing the recording signals at positions without microphones. To ensure stability, the rigid sphere model was used to define a surface pressure variation function that relates the pressure at two arbitrary points on the rigid spherical baffle. The proposal constitutes a preprocessing stage intended to be applied before array signal encoding or any other processing is performed.

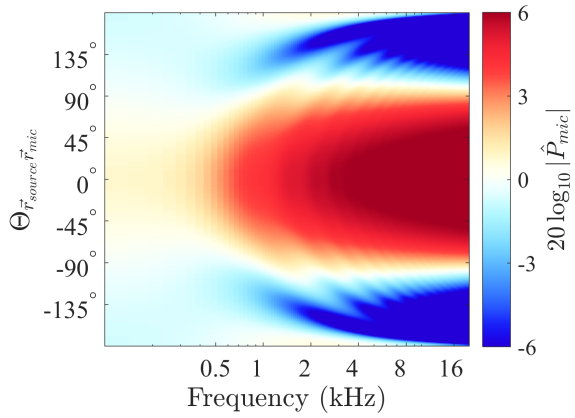
The remainder of this paper is organized as follows. Section 2 overviews the physical model of the rigid sphere. Section 3 presents the surface pressure variation function. Section 4 describes the method for enhancing the spatial resolution of rigid spherical microphone array recordings using the surface pressure variation function. Section 5 presents numerical examples with a rigid sphere. Concluding remarks are given in section 6.

2. **The acoustically rigid sphere model.**

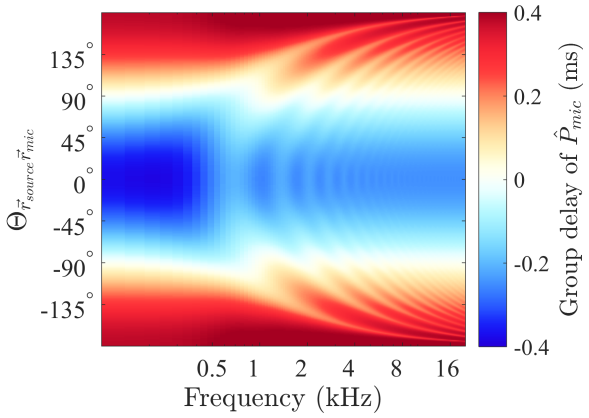
The interactions of sound pressure waves with an acoustically rigid sphere have been studied analytically in a number of theoretical acoustics texts [29, 30, 31]. The rigid sphere model has also been used in many design theories for spherical microphone arrays [14, 24, 25]. In addition, this model has been



(a) Geometry.



(b) Magnitude.



(c) Group delay.

FIGURE 1. Pressure on the surface of a rigid sphere of radius 8.5 cm due to a sound source at 150 cm from the center. Pressure is shown normalized by the free-field pressure measured at the center when the rigid sphere was not present.

validated through simulations and measurements in anechoic conditions [32], and through simulations in reverberant conditions [33].

Consider a rigid sphere of radius r_{mic} and a sound source at a distance $r_{\text{source}} > r_{\text{mic}}$ measured from the center of the rigid sphere. The total pressure generated by a sound source placed at a position $\vec{r}_{\text{source}}$ and measured by an ideal microphone placed at a position $\vec{r}_{\text{mic}}$ on the surface of the rigid sphere is analytically defined by the following expression [29, 30, 31]:

$$P_{\text{mic}} = P(\vec{r}_{\text{source}}, \vec{r}_{\text{mic}}, k) = -\frac{1}{kr_{\text{mic}}^2} \sum_{n=0}^{\infty} \frac{h_n(kr_{\text{source}})}{h_n'(kr_{\text{mic}})}(2n+1)L_n(\cos\Theta_{\vec{r}_{\text{source}}\vec{r}_{\text{mic}}}), \quad (1)$$

where

$$k = \frac{2\pi f}{c}. \quad (2)$$

Here, f denotes the frequency and c denotes the speed of sound in air. In (1), h_n denotes the spherical Hankel function of order n and the symbol $'$ indicates derivative with respect to the argument. In addition, L_n denotes the Legendre polynomial of order n evaluated at the cosine of the angle $\Theta_{\vec{r}_{\text{source}}\vec{r}_{\text{mic}}}$ between $\vec{r}_{\text{source}}$ and $\vec{r}_{\text{mic}}$.

Figure 1 illustrates the behavior of the rigid sphere model in (1) for a source distance $r_{\text{source}} = 150$ cm and a rigid baffle of radius $r_{\text{mic}} = 8.5$ cm. Pressure is shown to be

normalized by the free-field pressure P_0^{ff} measured at the center of the array when the rigid sphere is not present. Normalization was performed based on the following expression:

$$\hat{P}_{\text{mic}} = \frac{P_{\text{mic}}}{P_0^{\text{ff}}}, \tag{3}$$

where

$$P_0^{\text{ff}} = P^{\text{ff}}(\vec{r}_{\text{source}}, \vec{0}, k) = \frac{e^{jkr_{\text{source}}}}{r_{\text{source}}}, \tag{4}$$

and $j = \sqrt{-1}$. The hat symbol $\hat{}$ is used to denote free-field normalization hereafter.

Figure 1(a) shows the geometry describing the rigid sphere model. The colorbar in Fig 1(b) indicates the magnitude of the normalized model. The colorbar of Fig. 1(c) indicates the phase of the normalized model expressed in terms of the group delay. The ordinates in Figs 1(b) and 1(c) correspond to the angles $\Theta_{\vec{r}_{\text{source}}\vec{r}_{\text{mic}}}$ between the source and measurement points along a full circle on the sphere, whereas the abscissas indicate the audible frequency range in logarithmic scale. The angle 0° indicates the front measurement position directly facing the source, whereas the angles ±180° indicate the back position on the opposite side of the source. Owing to the rotational symmetry with respect to the line connecting the source position and the center of the array, these illustrations are representative of all possible positions over the rigid sphere.

In Figs. 1(b) and 1(c), constructive acoustic scattering on the side of the rigid sphere facing the source is reflected into higher magnitudes and earlier arrival times at angles $\Theta_{\vec{r}_{\text{source}}\vec{r}_{\text{mic}}} \in [-90°, 90°]$ when compared to the free-field pressure at the center P_0^{ff}. On the opposite side, at angles $\Theta_{\vec{r}_{\text{source}}\vec{r}_{\text{mic}}} \in [-180°, -90°] \cup [90°, 180°]$, the acoustic shadow of the rigid sphere produces smaller magnitudes and higher delays.

3. **Surface pressure variation function for the rigid sphere.** The model in (1) is used in this section to relate the pressure at two arbitrary points on the rigid sphere when a reference sound source position is assumed. This defines the following surface pressure variation function:

$$F_{1\mapsto 2} = \frac{P_2}{P_1}, \tag{5}$$

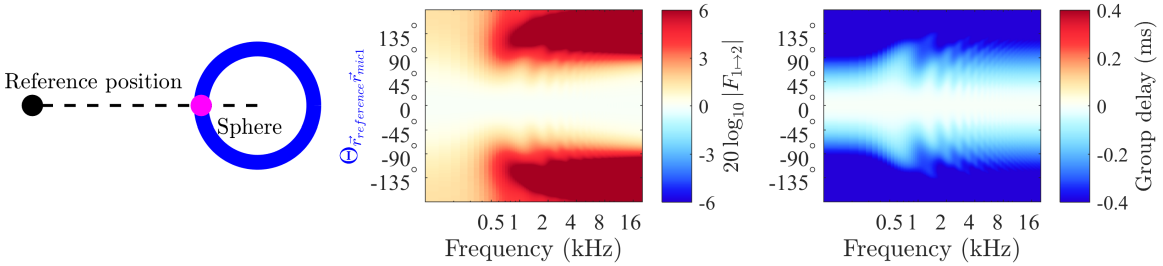
where

$$P_1 = P(\vec{r}_{\text{reference}}, \vec{r}_{\text{mic1}}, k), \text{ and } P_2 = P(\vec{r}_{\text{reference}}, \vec{r}_{\text{mic2}}, k). \tag{6}$$

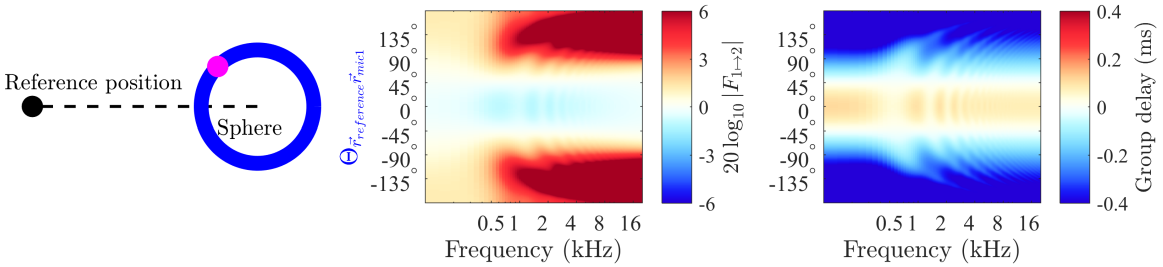
When a reference sound source position $\vec{r}_{\text{reference}}$ is assumed to be known, the function $F_{1\mapsto 2}$ represents the transmission of sound on the surface of the rigid sphere from one arbitrary measurement point $\vec{r}_{\text{mic1}}$ to another arbitrary measurement point $\vec{r}_{\text{mic2}}$.

Figure 2 shows the behavior of $F_{1\mapsto 2}$ for five illustrative cases. The reference sound source distance was set to $r_{\text{reference}} = 150$ cm, the rigid sphere radius was set to $r_{\text{mic}} = 8.5$ cm, the measurement point $\vec{r}_{\text{mic1}}$ is taken along a full circle on the sphere, and the measurement point $\vec{r}_{\text{mic2}}$ is fixed at one of five angles $0°, 45°, 90°, 135°$, and 180° at a time. These five angles are measured with respect to the reference position.

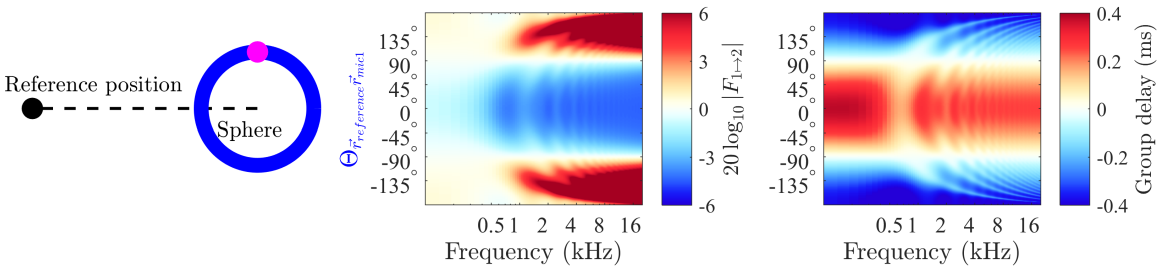
Figure 2(a) shows the case when $\vec{r}_{\text{mic2}}$ faces the reference point and is exactly aligned with it (i.e., the angle at which P_2 was calculated corresponds to 0° in the ordinate). Bounded variations of magnitude and group delay are observed within a wide and symmetric vicinity around 0°, which ranges from −45° to 45°. In Figs. 2(b) to 2(e), it is observed that the vicinity of bounded variations around $\vec{r}_{\text{mic2}}$ (i.e., around the angle in the ordinates at which P_2 is calculated) loses extension as $\vec{r}_{\text{mic2}}$ moves towards the opposite side. The vicinity of bounded variations around $\vec{r}_{\text{mic2}}$ also loses symmetry as $\vec{r}_{\text{mic2}}$ approaches 90°.



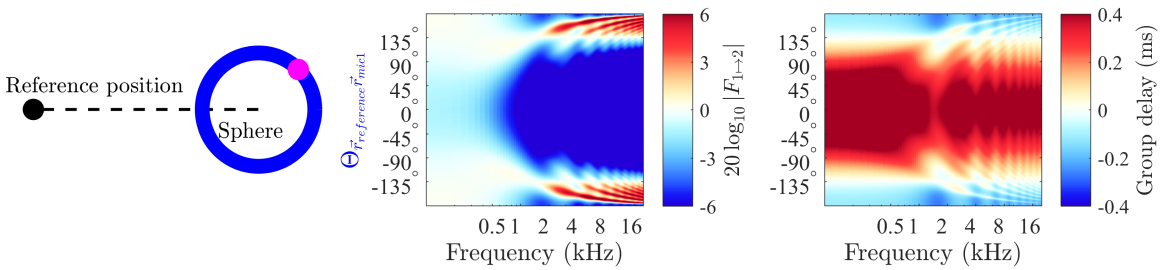
(a) $\vec{r}_{\text{mic2}}$ at 0°.



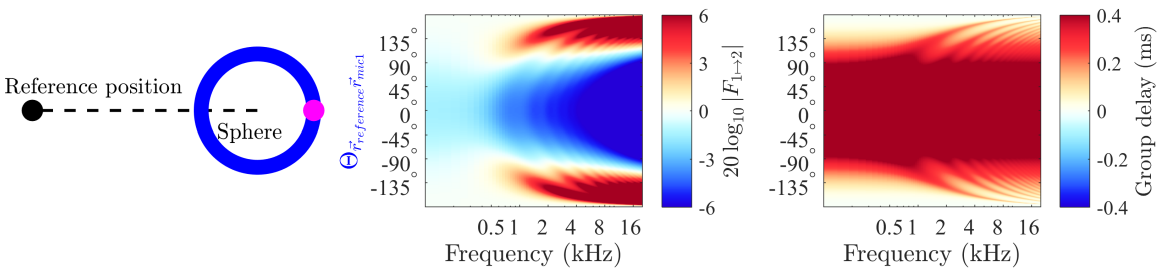
(b) $\vec{r}_{\text{mic2}}$ at 45°.



(c) $\vec{r}_{\text{mic2}}$ at 90°.



(d) $\vec{r}_{\text{mic2}}$ at 135°.



(e) $\vec{r}_{\text{mic2}}$ at 180°.

FIGURE 2. Examples of surface pressure variation functions $F_{1\mapsto2}$ in (5).

The illustrations in Fig. 2 show the potential of $F_{1\mapsto 2}$ for predicting sound pressure signals in the vicinity of a microphone. The size of the vicinity is smaller for microphones placed on the opposite side of the source. This potential motivates the definition of a method to generate virtual microphone signals to enhance the spatial resolution of spherical array recordings.

4. **Virtual microphone generation method for spatial resolution enhancement.** The surface pressure variation function $F_{1\mapsto 2}$ in (5) is used in this section to define a virtual microphone generation method for the spatial resolution enhancement of rigid spherical microphone array recordings.

The function $F_{1\mapsto 2}$ is interpreted here as a surface pressure interpolation filter on the rigid sphere that is applied to the microphone recording signals. Two cases are taken into consideration for the generation of virtual microphone signals in the vicinity of real microphones in the array. In the first case, the vicinity is considered around a single real microphone. In the second case, the vicinity is considered around the two nearest real microphones. Each case is formulated below.

4.1. **Surface pressure interpolation from a single nearest microphone.** S_1 denotes the signal recorded by a real microphone placed at a position $\vec{r}_{\text{mic1}}$ in the array. S_2 denotes the virtual microphone signal to be generated at a position $\vec{r}_{\text{mic2}}$ by applying the filter $F_{1\mapsto 2}$ to S_1 as follows:

$$S_2 = F_{1\mapsto 2} S_1. \tag{7}$$

Note that the application of (7) requires the specification of two positions to calculate $F_{1\mapsto 2}$ in accordance to (5). These are the reference position $\vec{r}_{\text{reference}}$ and the virtual microphone position $\vec{r}_{\text{mic2}}$. The position $\vec{r}_{\text{mic1}}$ of the real microphone from which interpolation is performed can be simply selected by a nearest neighbor search. This search consists of two steps. First, the angles between $\vec{r}_{\text{mic2}}$ and each real microphone in the array are calculated. Next, the microphone in the array that creates the smallest angle from $\vec{r}_{\text{mic2}}$ is selected.

4.2. **Surface pressure interpolation from two nearest microphones.** S_{1a} and S_{1b} denote the signals recorded by two real microphones placed at positions $\vec{r}_{\text{mic1a}}$ and $\vec{r}_{\text{mic1b}}$ in the array, respectively. S_{2a} and S_{2b} denote the virtual signals to be generated at a same position $\vec{r}_{\text{mic2}}$ by correspondingly applying the filters $F_{1a\mapsto 2}$ and $F_{1b\mapsto 2}$ in such a way that

$$S_{2a} = F_{1a\mapsto 2} S_{1a}, \tag{8}$$

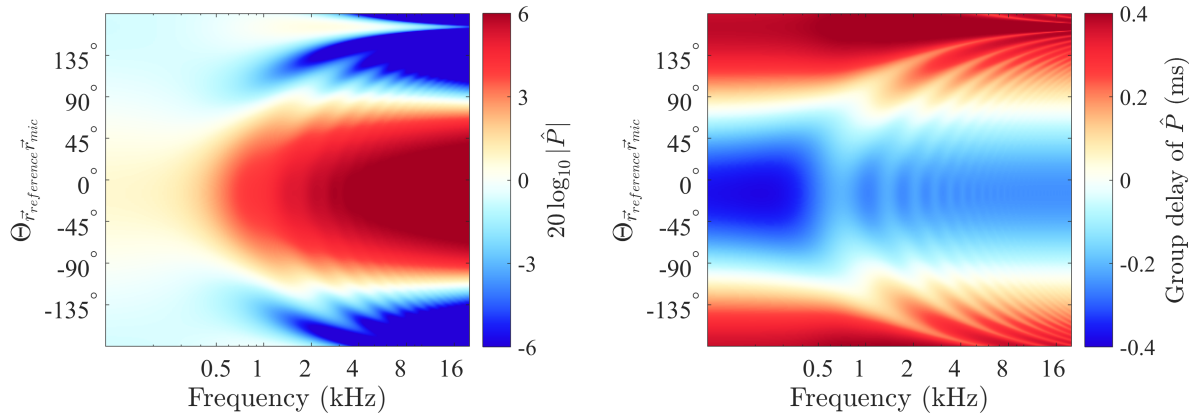
$$S_{2b} = F_{1b\mapsto 2} S_{1b}. \tag{9}$$

The virtual signals S_{2a} and S_{2b} are finally combined into a single signal (corresponding to a virtual microphone at $\vec{r}_{\text{mic2}}$) by averaging based on the geometric mean as follows:

$$S_2 = (S_{2a} S_{2b})^{\frac{1}{2}} = (|S_{2a}||S_{2b}|)^{\frac{1}{2}}\, e^{j\frac{\angle S_{2a} + \angle S_{2b}}{2}}. \tag{10}$$

The combination of virtual signals based on the geometric mean has the advantage of considering the proper application of gains and delays in the frequency domain. The resulting magnitude corresponds to the geometric mean of magnitudes, whereas the resulting phase corresponds to the arithmetic average of phases. Another advantage of using the geometric mean is that the general combination of more than two microphone signals is straightforward. Note that if the whole combination was performed based on a simple arithmetic average, the results would not properly preserve the superposition of pressures.

Note also that the application of (8) to (10) requires the specification of a reference position $\vec{r}_{\text{reference}}$ and a virtual microphone position $\vec{r}_{\text{mic2}}$. The real microphone positions



(a) Magnitude. (b) Group delay.

FIGURE 3. Target microphone signals.

$\vec{r}_{\text{mic1a}}$ and $\vec{r}_{\text{mic1b}}$ can be simply selected by a nearest neighbor search. This search relies on selecting the two microphones in the array that form the smallest angles from $\vec{r}_{\text{mic2}}$.

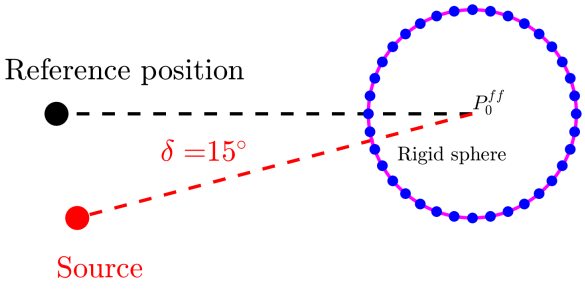
5. **Numerical examples.** This section presents illustrative results based on the numerical evaluation of the virtual microphone generation method, considering the cases described in sections 4.1 and 4.2.

The numerical evaluations considered the effects of a mismatch between the position of the sound source to be recorded and the reference position used to calculate the surface pressure interpolation filters $F_{1\mapsto 2}$. The mismatch was considered only along the angle and it was represented by a deviation angle δ. In addition, the numerical evaluations also considered the effects of having a finite number of real microphones distributed in a circle around the array. Given the rotational symmetry of the rigid sphere model, the results presented in this section constitute representative cases for the whole sphere.

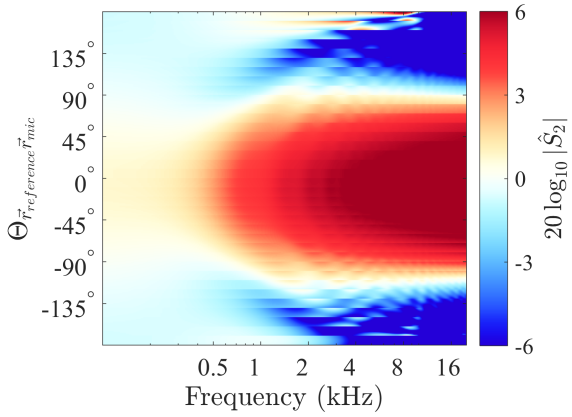
Figure 3 shows the target microphone signals that were used as reference to examine the generated virtual microphone signals. The target signals were simulated for a sound source (to be recorded) at a distance $r_{\text{source}} = 150$ cm, and 360 microphone positions equiangularly distributed around a circle on a rigid sphere of radius $r_{\text{mic}} = 8.5$ cm. The ordinate corresponds to the angle between the target microphone position and the reference position used to calculate the interpolation filters $F_{1\mapsto 2}$ in (5). Note the angular deviation of 15° when compared with Fig. 1. The hat symbol $\hat{}$ in the colorbars is similarly used hereafter to denote free-field normalization based on (3).

In all numerical experiments, the interpolation filters $F_{1\mapsto 2}$ were calculated for a reference angle 0° and for a reference distance $r_{\text{reference}} = 150$ cm, and the sound source to be recorded was placed at an angle deviation $\delta = 15°$ from the reference position, as shown later in Figs. 4(a) and 5(a).

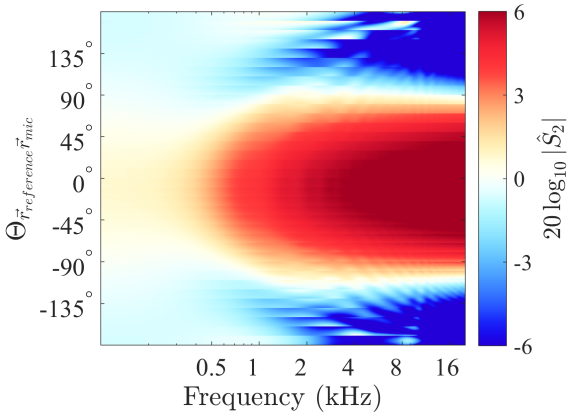
Figure 4 shows the results obtained when the number of real microphones was 36, which correspond to an angular separation of 10° between two microphones. Figures 4(b) and 4(d) show that interpolation from a single nearest microphone based on (7) yields accurate results in terms of the magnitude and group delay when the virtual microphone signals were generated on the side of the array facing the source. On the opposite side, however, the patterns corresponding to the acoustic shadow of the sphere were not properly replicated at middle and higher frequencies. On the other hand, Figs. 4(c) and 4(e) show that the generation of virtual signals on the opposite side can be improved by including the two nearest microphones in the interpolation based on (8) to (10). Note that all virtual



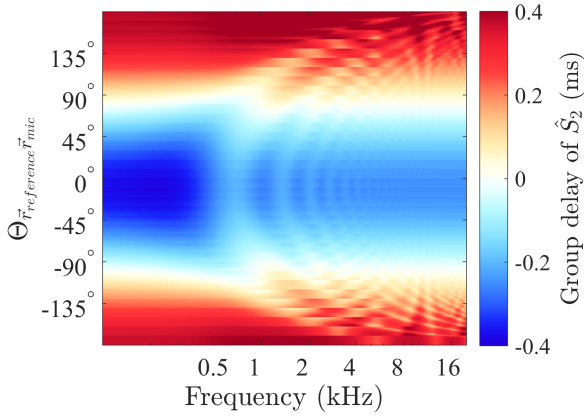
(a) Geometry.



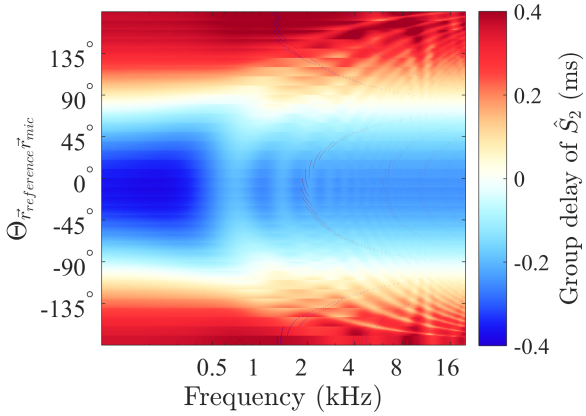
(b) Interpolation from one nearest microphone.



(c) Interpolation from two nearest microphones.



(d) Interpolation from one nearest microphone.

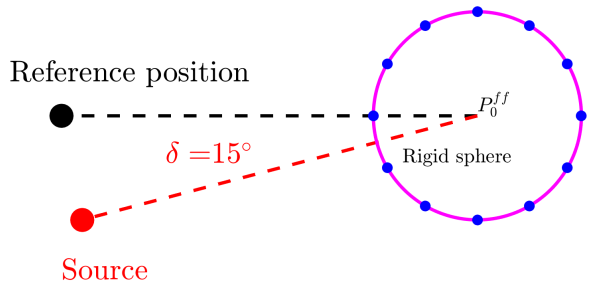


(e) Interpolation from two nearest microphones.

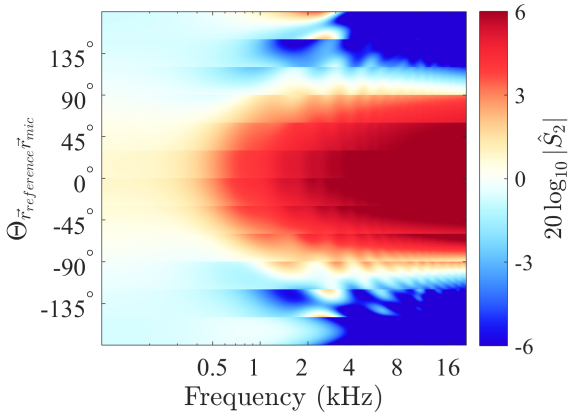
FIGURE 4. Virtual microphone signals obtained by surface pressure interpolation. The initial data corresponds to 36 microphones equiangularly distributed (10°) along a circle on the surface of the rigid sphere. The recorded sound source was deviated along the direction by an angle of $\delta = 15°$ from the reference position. Top panels (b) and (c) show magnitudes of interpolated signals, whereas bottom panels (d) and (e) show group delays of interpolated signals.

signals lacked discontinuities along the angle because of microphone selection based on the nearest neighbor search.

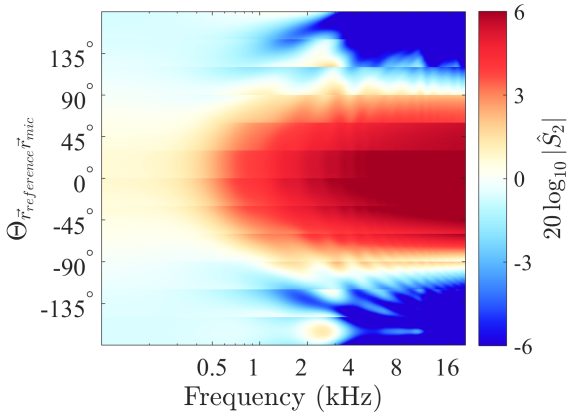
Figure 5 shows the results obtained when the number of real microphones was 12 (microphone angular separation of 30°). Figures 5(b) and 5(d) show that interpolation from



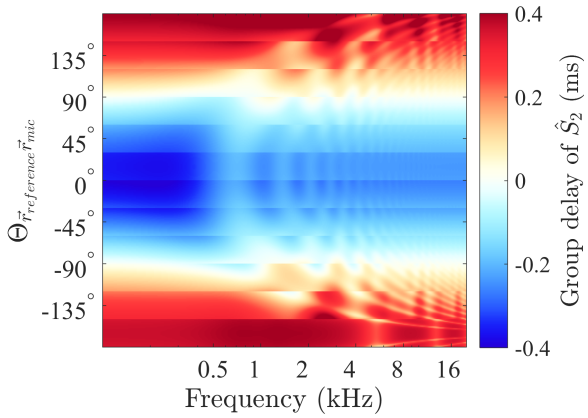
(a) Geometry.



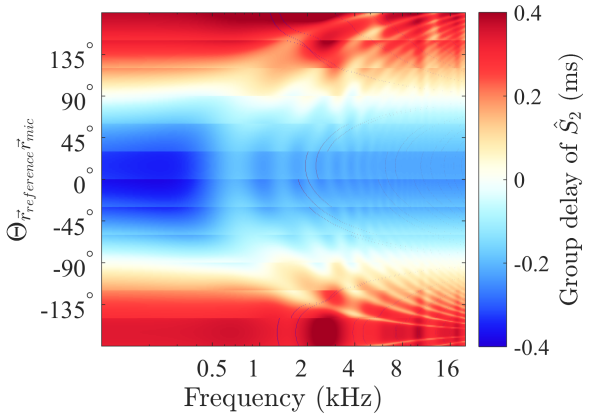
(b) Interpolation from one nearest microphone.



(c) Interpolation from two nearest microphones.



(d) Interpolation from one nearest microphone.



(e) Interpolation from two nearest microphones.

FIGURE 5. Virtual microphone signals obtained by surface pressure interpolation. The initial data corresponds to 12 microphones equiangularly distributed (30°) along a circle on the surface of the rigid sphere. The recorded sound source was deviated along the direction by an angle of $\delta = 15°$ from the reference position. Top panels (b) and (c) show magnitudes of interpolated signals, whereas bottom panels (d) and (e) show group delays of interpolated signals.

a single nearest microphone based on (7) yields acceptable results in terms of magnitude and group delay, when the virtual microphone signals were generated on the side of the array facing the source. On the opposite side, however, the patterns corresponding to the acoustic shadow of the sphere were not properly replicated at the middle and higher

frequencies. On the other hand, Figs. 5(c) and 5(e) show that the generation of virtual signals on the opposite side can be improved by including the two nearest microphones in the interpolation based on (8) to (10). In all cases, virtual signals also lacked discontinuities along the angle because of microphone selection based on the nearest neighbor search.

Of special interest is the objective examination of the effects that the deviation angle δ would produce on the performance of interpolation. Two additional numerical experiments are presented below to investigate these phenomena in both anechoic and reverberant conditions.

Objective evaluation is based on the following interpolation error:

$$E(\Theta_{\vec{r}_{\text{reference}}\vec{r}_{\text{mic}}}, k) = \frac{S(\Theta_{\vec{r}_{\text{reference}}\vec{r}_{\text{mic}}}, k)}{P(\Theta_{\vec{r}_{\text{reference}}\vec{r}_{\text{mic}}}, k)}, \tag{11}$$

where S and P denote the interpolated and target signals, respectively. To examine the spatiospectral distortion, the magnitude of E was first expressed in logarithmic scale, and then the RMS of the result was calculated along $\Theta_{\vec{r}_{\text{reference}}\vec{r}_{\text{mic}}}$ and k. This procedure was repeated for several deviation angles δ.

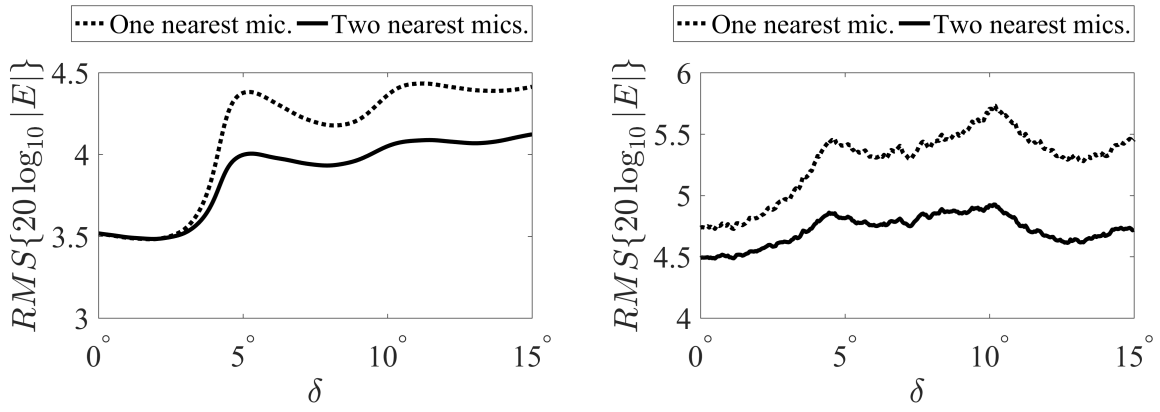
Recording in a reverberant condition was simulated using the algorithm in [33] to describe a rigid sphere of $r_{\text{mic}} = 8.5$ cm radius placed inside a rectangular parallelepiped room. The image method was used to describe a first-order reflection model of the room. The center of the coordinates coincides with a bottom corner of the room. The dimensions of the room included a 5 m width (along x), 6 m length (along y), and 4 m height (along z). The reflection coefficients of all walls were 0.3. The center position of the microphone array was $(1.6, 4.05, 1.7)$ m. In addition, the initial position of the sound to be recorded was $(3.3, 4.05, 1.7)$ m; during the repetitions of numerical experiments, this source was deviated by an angle δ toward the positive $y-$axis.

Fig. 6 shows the interpolation error as a function of the deviation angle δ. Fig. 6(a) shows that in anechoic conditions, both interpolation methods yield similar results when the deviation angle is small. However, the advantage of relying on two nearest microphones is observed at greater deviation angles. In reverberant conditions, Fig. 6(b) shows that interpolation from two nearest microphones outperforms that from one nearest microphone. Nevertheless, the relation between the RMS of interpolation errors and the deviation angle is not simple.

6. **Conclusions.** A method to generate virtual microphone signals for the spatial resolution enhancement of rigid spherical microphone arrays was presented and evaluated. This method requires prior knowledge about source positions and can be used in recording environments such as conference rooms or concert halls where the source positions are often confined to a small region of space. Using this prior knowledge, a surface pressure variation function for the rigid sphere was introduced relying on a physical model. On the basis of this function, the spatial resolution enhancement method for spherical arrays was formulated.

The proposed method adds virtual microphones to the array by synthesizing recording signals at positions without microphones. The proposal constitutes a preprocessing stage intended to be applied before array signal encoding.

Numerical experiments with a rigid sphere show that the enhancement of spatial resolution is possible all over the sphere if the number of real microphones is sufficiently large. On the side of the array that is opposite to the sources, the performance decreases with increasing frequency. This is overcome based on interpolation from plural microphones, which significantly improves the performance on the opposite side.



(a) Anechoic condition.

(b) Reverberant condition.

FIGURE 6. Overall interpolation error. The initial data corresponds to 12 microphones equiangularly distributed (30°) along a circle on the surface of the rigid sphere. The position of the sound source to be recorded was deviated from the reference position along an angle δ with values from 0° to 15°. The reverberant condition was simulated with a first-order reflection model of a rectangular parallelepiped room of 5 m width, 6 m length, and 4 m height. The reflection coefficients of all walls were set to 0.3.

Extensions of this work can include improvements in phase unwrapping algorithms, surface pressure interpolation from more than the two nearest microphones, and further considerations of surface pressure interpolation under diffuse fields.

Acknowledgment. A part of this study was supported by a JSPS Grant-in-Aid for Scientific Research (nos. 24240016 and 16H01736) and the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet."

REFERENCES

[1] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence: Teleop. Virt. Env.*, vol. 5, no. 3, pp. 290–301, Jan. 1996.

[2] M. Kobayashi, K. Ueno, and S. Ise, "The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study," *Presence: Teleop. Virt. Env.*, vol. 24, no. 2, pp. 163–174, May 2015.

[3] D. Vastfjall, "The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments," *Cyberpsychol. Behav.*, vol. 6, no. 2, pp. 181–188, 2003.

[4] 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," *Interdiscipl. Inform. Sci.*, vol. 18, no. 2, pp. 71–82, 2012.

[5] S. Sakamoto, Y. Osada, Y. Suzuki, and J. Gyoba, "The effects of linearly moving sound images on self-motion perception," *Acoust. Sci. Technol.*, vol. 25, no. 1, pp. 100–102, 2004.

[6] B. E. Riecke, D. Feuereissen, J. J. Rieser, and T. P. McNamara, "Spatialized sound enhances biomechanically-induced self-motion illusion (vection)," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2799–2802.

[7] W. Teramoto, S. Sakamoto, F. Furune, J. Gyoba, and Y. Suzuki, "Compression of auditory space during forward self-motion," *PLoS ONE*, vol. 7, no. 6, p. e39402, Jun. 2012.

[8] B. Keshavarz, L. J. Hettinger, D. Vena, and J. L. Campos, "Combined effects of auditory and visual cues on the perception of vection," *Exp. Brain. Res.*, vol. 232, no. 3, pp. 827–836, 2014.

[9] B. E. Riecke and J. Schulte-Pelkum, "An integrative approach to presence and self-motion perception research," in *Immersed in Media: Telepresence Theory, Measurement & Technology*, M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, Eds. Cham: Springer International Publishing, 2015, pp. 187–235.

[10] A. Steed and R. Schroeder, "Collaboration in immersive and non-immersive virtual environments," in *Immersed in Media: Telepresence Theory, Measurement & Technology*, M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, Eds. Cham: Springer International Publishing, 2015, pp. 263–282.

[11] V. Algazi and R. Duda, "Headphone-based spatial sound," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 33–42, Jan. 2011.

[12] J. Blauert and R. Rabenstein, "Providing surround sound with loudspeakers: A synopsis of current methods," *Arch. Acoust.*, vol. 37, no. 1, pp. 5–18, 2013.

[13] T. Betlehem, W. Zhang, M. Poletti, and T. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.

[14] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets," *Acoust. Sci. Technol.*, vol. 38, no. 2, pp. 51–62, Mar. 2017.

[15] J. Eargle, *The Microphone Book: From mono to stereo to surround-a guide to microphone design and application.* CRC Press, 2012.

[16] F. Rumsey, *Spatial Audio.* CRC Press, 2012.

[17] V. R. Algazi, R. O. Duda, and D. M. Thompson, "Motion-tracked binaural sound," *J. Audio Eng. Soc.*, vol. 52, no. 11, pp. 1142–1156, 2004.

[18] 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 119*, New York, USA, Oct. 2005.

[19] C. D. Salvador, S. Sakamoto, J. Trevino, 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.

[20] S. Sakamoto, S. Hongo, and Y. Suzuki, "3d sound-space sensing method based on numerous symmetrically arranged microphones," *IEICE Trans. Fundamentals*, vol. E97-A, no. 9, pp. 1893–1901, Sep. 2014.

[21] 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. Technol.*, vol. 36, no. 6, pp. 516–526, 2015.

[22] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Numerical evaluation of binaural synthesis from rigid spherical microphone array recordings," in *Proc. AES Int. Conf. Headphone Technology*, Aalborg, Denmark, Aug. 2016.

[23] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Spatial accuracy of binaural synthesis from rigid spherical microphone array recordings," *Acoust. Sci. Technol.*, vol. 38, no. 1, pp. 23–30, Jan. 2017.

[24] 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 2002, pp. 1781–1784.

[25] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech, Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[26] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *23rd Int. Conf. Audio Eng. Soc.*, May 2003.

[27] J. Trevino, T. Okamoto, Y. Iwaya, and Y. Suzuki, "High order ambisonic decoding method for irregular loudspeaker arrays," in *Proc. 20th Int. Congress Acoust.*, 2010, pp. 23–27.

[28] J. Trevino, T. Okamoto, C. D. Salvador, Y. Iwaya, Z. Cui, S. Sakamoto, and Y. Suzuki, "High-order ambisonics auditory displays for the scalable presentation of immersive 3d audio-visual contents," in *Proc. 23rd Int. Conf. Artificial Reality and Telexistence*, Tokyo, Japan, 2013.

[29] P. M. Morse and K. U. Ingard, *Theoretical Acoustics.* Princeton University Press, 1987.

[30] J. J. Bowman, T. B. A. Senior, and P. Uslenghi, *Electromagnetic and acoustic scattering by simple shapes.* New York, NY, USA: Hemisphere, 1987.

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

[32] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.

[33] D. P. Jarret, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.