Effects of spatial aliasing in sound field reproduction:
Reproducibility of binaural signals

Makoto Otani1,*; Hikaru Watabe2; Takao Tsuchiya3 and Yukio Iwaya4

1Graduate School of Engineering, Kyoto University,
Kyoto daigaku-katsura, Nishikyo-ku, Kyoto, 615–8540 Japan
2Graduate School of Science and Technology, Shinshu University,
4–17–1 Wakasato, Nagano, 380–8553 Japan
3Faculty of Science and Engineering, Doshisha University,
1–3 Tatara Miyakodani, Kyotanabe, 610–0394 Japan
4Faculty of Engineering, Tohoku Gakuin University,
1–13–1 Chujo, Tagajo, 985–8537 Japan

(Received 24 August 2016, Accepted for publication 22 December 2016)

Abstract: To realize physically accurate sound field reproduction, the boundary surface of a sound field to be reproduced should be spatially discretized with intervals smaller than a half wavelength. Otherwise, spatial aliasing will occur in the reproduced field, which leads to low physical reproducibility. Therefore, accurate sound field reproduction covering the full audible range up to approximately 20 kHz requires an impractically large number of sampling points, namely, the number of microphones and loudspeakers. However, it may be possible to reduce the number of sampling points if the degradation in the physical performance due to spatial aliasing does not degrade the spatial perception of the reproduced sound field. To achieve such perceptual optimization of a sound field reproduction system, it should be clarified how spatial aliasing has negative effects on the physical and perceptual reproducibility of a sound field reproduction system. Therefore, as a first step, to investigate the physical reproducibility of sound field reproduction with spatial aliasing, we numerically simulate the reproduced sound field and binaural signals that will be reproduced when a listener is inside the reproduced sound field. The numerical results of the reproduced sound field with spatial aliasing showed that sampling intervals larger than a half wavelength yield unnecessary wave fronts that reach a listener 1 ms after the main wave fronts. Furthermore, the results of the numerical simulation of a binaural signal, employing the boundary element simulation of a human head, suggested that interaural time differences and level differences are approximately reproduced when the upper-bound frequency of physically accurate reproduction is greater than 4 and 8 kHz, respectively.

Keywords: Sound field reproduction, Spatial aliasing, Boundary element simulation, Binaural signal

PACS number: 43.60.Tj, 43.20.Bi [doi:10.1250/ast.38.147]

1. INTRODUCTION

For the realization of sound field reproduction followed by the presentation of a virtual auditory space to a listener, several approaches and methods have been proposed. Firstly, Camras proposed an approach of sound field reproduction based on Huygens’ principle in 1968 [1]. Subsequently, Berkhout et al. proposed wave field synthesis (WFS) [2] based on the Rayleigh integral and Ise proposed a principle of boundary surface control (BoSC) [3] employing the Kirchhoff–Helmholtz integral equation and inverse systems, both of which give mathematical backgrounds to sound field reproduction. Theoretically, these methods can reproduce a primary sound field in a secondary field by locating monopole or dipole sources on the boundary surface of the reproduced volume or by controlling the sound pressures and their gradients on the boundary surface using the inverse system.

However, because, considering the spatial sampling, the spatial interval of sampling points on a boundary surface must not be greater than λ/2 where λ is the wavelength, physically accurate sound field reproduction requires a large-scale system consisting of a very large number of sampling points, i.e., microphones and loud-
speakers, especially for high frequencies. Therefore, it is practically difficult to realize sound field reproduction that is physically accurate for the full audible frequency range of humans with currently available hardware including loudspeakers and digital-to-analogue converters. When employing a spatial sampling interval larger than \( \lambda/2 \), the resulting reproduced sound field is distorted owing to spatial aliasing.

Despite such hardware limitations, the demand for highly realistic sound field reproduction is increasing but very little is known about the effects of spatial aliasing on the reproducibility of a sound field. To construct a practically realistic sound field reproduction system with the hardware limitations, it is important to clarify, from both the physical and perceptual viewpoints, how spatial aliasing negatively affects the reproduced sound field and the extent to which it does. The authors previously reported numerical simulation results of sound field reproduction with various spatial sampling intervals, showing that insufficient spatial sampling leads to deformation of the main wave front and the production of unnecessary wave fronts; these negative effects of spatial aliasing are more prominent with larger sampling intervals. In this paper, the authors numerically calculated by combining the boundary elements of the simulated binaural signals to be observed in a reproduced sound field. Furthermore, based on the simulated binaural signals in the reproduced field, the effects of spatial aliasing on binaural inputs are discussed from the viewpoint of spatial perception.

The following sections are organized as below. Section 2 introduces the theoretical background of sound field reproduction. Section 3 briefly reviews the results of the numerical simulation of a reproduced sound field with various spatial sampling intervals. Section 4 presents the method and results of a binaural signal simulation in a reproduced sound field. Sections 5 and 6 respectively provide a discussion and conclusions.

2. THEORIES OF SOUND FIELD REPRODUCTION

Sound field reproduction is based on integral equations derived from the wave equation. Berkhout et al. proposed WFS [2] based on the Rayleigh integral. Assuming an infinite plane \( z = z_1 \) as a boundary surface \( S \), the Kirchhoff–Helmholtz integral equation (KHIE) can be transformed to Rayleigh integral II, written as

\[
p(s) = \frac{|z - z_1|}{2\pi} \int_S p(r) \frac{\partial G(r,s)}{\partial n} \frac{e^{-jkr}}{|r - s|} \, ds,
\]

where \( s \) is a point in the sound field, \( r \) is a point on the boundary surface, \( p \) is the sound pressure, \( j \) is the imaginary unit, and \( k \) is the wave number. This equation means that the sound field \( p(s) \) is reproduced in the secondary field by recording sound pressures \( p(r) \) on \( S \) in the primary field and, subsequently, by locating the dipole sources of the source strength \( p(r) \) on \( S \) in the secondary field.

Ise proposed the principle of BoSC [3] based on the KHIE, written as

\[
p(s) = \int_S \{ p(r) \frac{\partial G(r,s)}{\partial n} - \frac{\partial p(r)}{\partial n} G(r,s) \} \, ds,
\]

where \( G(r,s) \) is the Green’s function. In BoSC, it is assumed that \( G(r,s) \) and its normal derivative \( \partial G(r,s)/\partial n \) are constants characterized by the boundary surface geometry; \( p(r) \) and \( \partial p(r)/\partial n \) are reproduced on the boundary surface in the secondary field.

Recently, Kimura and Kakehi proposed an alternative formulation for sound field reproduction based on the following integral equation [5]:

\[
p(s) = \int_S \int_S jk(1 - \cos \theta_s)p(r)G(r,s)dS
\]

\[
\cos \theta_s = \frac{(r - r_s) \cdot n}{|r - r_s||n|},
\]

which is derived from the Fresnel–Kirchhoff diffraction formula by assuming a spherical boundary surface and a receiver \( s \) located at the center of the sphere. Here, \( r_s \) denotes the source position in the primary field, and \( \theta_s \) is the angle between \( r - r_s \) and the surface normal vector \( n \) at \( r \) on the surface. According to Eq. (1), the sound field inside the boundary can be reproduced by locating monopole sources on the boundary surface having a strength of \( jk(1 - \cos \theta_s)p(r) \), namely, sound pressures recorded in the primary field with cardioid directivity whose main axis corresponds to the surface normal.

The current work employs the method of Kimura and Kakehi because it enables the reproduction of sound waves from any direction by using only monopole sources, although it can handle only a spherical boundary. Discretization of the boundary surface \( S \) transforms Eq. (1) to

\[
p(s) = \sum_{i=1}^{M} jk(1 - \cos \theta_{si})p(r_i)G(r_i,s)\Delta S_i
\]

\[
\cos \theta_{si} = \frac{(r_i - r_s) \cdot n_i}{|r_i - r_s||n_i|},
\]

where \( \Delta S_i \) is the surface area of an element including \( r_i \) and \( n_i \) is the surface normal at \( r_i \). Therefore, the sound pressure field \( p(s) \) is reproduced by recording sound pressures with cardioid directivity at each discretized point \( r_i \) and by locating monopole sources radiating the recorded signals at the corresponding discretized points in the secondary field.
3. Reproducibility of Sound Field

This section reviews the results of the authors’ previous work [4], demonstrating the reproducibility of sound field reproduction with various spatial sampling intervals.

3.1. Sampling Point Configuration

To numerically examine the effects of spatial aliasing on the physical reproducibility of a sound field, five sets of sampling points were prepared with various spatial sampling intervals, as illustrated in Fig. 1. The boundary of the control area was assumed to be a sphere of 0.25 m radius. Sampling points are distributed on its surface. The finest set consists of 11,312 sampling points whose maximum inter-control-point distance is approximately 12 mm and is theoretically valid up to 14 kHz. The other four sets consist of 2,928, 778, 178, and 44 sampling points. These sets are valid up to 8, 4, 2, and 1 kHz, respectively. Note that the distribution of the sampling points on the sphere surface is not perfectly uniform or symmetric. See [4] for the details of the generation of sampling points.

3.2. Numerical Conditions

Figure 2 demonstrates the configuration of a point source, control area, and observation area. The point source is located 2.346 m distant from the center of the control area, i.e., the center of the sphere. Sound pressures were numerically observed at two-dimensional 1 cm grid points in a 0.51 m square region (2,704 points in total) that covers the control area. Although the numerical simulation was performed in three-dimensional space, sound pressures were calculated in a plane that passes through the point source and the center of the sphere and is perpendicular to the z axis. This is because it is difficult to express a three-dimensional visualization of a sound field in an easily comprehensible and paper-based way. The plane on which the sound pressures were calculated corresponds to the horizontal plane that passes through both ears of the head model employed in Sect. 4, in which localization cues within the horizontal plane are discussed.

The numerical simulation of the reproduced sound field was performed as follows. First, impulse responses were calculated from the point source to sampling points with a cardioid directivity. Second, impulse responses from the sampling points to observation points were calculated. These impulse responses (sampling rate: 44.1 kHz, signal length: 512 points) were generated by applying an inverse Fourier transform to transfer functions obtained by the Green’s function,

\[ G = \frac{1}{4\pi|\vec{r}|} e^{-jk|\vec{r}|}, \]

where \(|\vec{r}|\) corresponds to the distance from the point source to each sampling point \(|\vec{r}_i - \vec{r}_s|\) or that from each sampling point to each observation point \(|\vec{s} - \vec{r}_i|\). The cardioid directivity is applied to the impulse responses by multiplying by a direction-dependent coefficient of \(1 - \cos \theta_{s,i}\).

Finally, the waveform observed at each observation point as a result of sound field reproduction was calculated by convolving a directive impulse response from the point source to each sampling point and the impulse response from each control point to the corresponding observation point, and subsequently summing them.

Because it is difficult to accurately estimate the surface area of each element \(\Delta S_i\) in Eq. (2), its value is...
approximately computed. Here, $\Delta S_i$ is roughly proportional to the square of the distance from $r_i$ to its closest neighboring point. Thus, letting $d_i$ be the distance from $r_i$ to its closest neighboring point yields $\Delta S_i \propto d_i^2$. Although this approximation does not provide an absolute value of $\Delta S_i$, the obtained approximation of a relative value of $\Delta S_i$ is sufficient for comparisons among different sets of sampling points. Note that, regardless of the limitation in the upper-bound frequency of each sampling-point set, the calculation of the waveform was performed for all the frequencies up to 22.05 kHz for the clarification of spatial aliasing effects in the entire audible range.

3.3. Results

Figure 3 illustrates the wave fronts generated in the reproduced sound field with each sampling-point configuration when the main wave front arrives at the center of the control area. The reproduced sound field is not symmetric about the $y$-$z$ plane despite the sound source being on the $y$-$z$ plane. This is because of the non-uniform and asymmetric distribution of the sampling points as mentioned in Sect. 3.1. Note that Fig. 3(a) depicts a wave front observed in the primary field that was simulated by the Green’s function from the point source to each observation point. Figure 3(a) shows that a spherical wave spreads from the point source. In contrast, Figs. 3(b)–3(f) demonstrate that sound field reproduction leads to a more distorted main wave front for a larger sampling interval. In particular, as shown in Figs. 3(e) and 3(f), sampling with 178 points (2 kHz) and 44 points (1 kHz) resulted in greatly distorted main wave fronts. Furthermore, in addition to such distortion, a larger sampling interval yields unwanted "after-effect" wave fronts that arrive from various directions after the arrival of the main wave front. On the other hand, in the cases of relatively fewer sampling points (Figs. 3(b)–3(d)), such degradation in the reproduced sound field is not so prominent.

Figures 4(a)–4(e) depict the spatial distribution of the signal-to-noise ratio (SNR) for all the sampling point configurations, calculated as

$$\text{SNR} = 10 \log_{10} \left( \frac{\sum_{n=1}^{m} (h_p[n])^2}{\sum_{n=1}^{m} (h_p[n] - h_s[n])^2} \right) [\text{dB}],$$

where $h_p[n]$ and $h_s[n]$ are the sound pressure in the time domain at sample $n$ for the primary and reproduced fields, respectively, and $m$ is the signal length. Figure 4 shows that the SNR decreases in the entire control area with increasing sampling interval.
4. REPRODUCIBILITY OF BINAURAL SIGNALS

As shown in the previous section, a larger sampling interval leads to lower reproducibility of the sound field because of spatial aliasing effects. However, it has not been revealed how these physical effects of spatial aliasing negatively affect human spatial perception such as sound image localization or spatial impression. Thus, to investigate the physical parameters that are most relevant to spatial perception, such as interaural time differences (ITDs), interaural level differences (ILDs), and spectral cues, a numerical simulation was performed to calculate the binaural signals that will be observed when a human head exists in a reproduced sound field.

Impulse responses from sampling points to both ears were calculated by applying the boundary element method (BEM) to a scattering field whose boundary consists of a human head [6]. Figure 5 depicts the boundary element model of the human head, which was generated from a magnetic resonance imaging scan and consists of approximately 28,000 triangular elements. This head model is valid up to 20 kHz when five elements per wavelength is assumed. The surface of the head model is assumed to be acoustically rigid. The front and top of the head correspond to the y and z axes, respectively, in the coordinate system shown in Fig. 2. The midpoint between both ears corresponds to the center of the spherical control region. Using the BEM and the head model, the impulse responses from all the sampling points to both ears are obtained. Binaural signals that will be observed in the reproduced sound field were calculated by convolving the directive impulse response from the point source to each sampling point with the impulse response from the corresponding sampling point to both ears and summing them over all the sampling points.

Figure 6 shows the waveforms of the resulting binaural signal at the left ear. The waveforms that will be observed in the primary field are presented by black lines and labeled as “Primary” in the figures. Hereafter, the upper-bound frequencies (14, 8, 4, 2, and 1 kHz) at which spatial aliasing does not occur are denoted as “14k,” “8k,” “4k,” “2k,” and “1k,” respectively. Figure 6 shows that the waveforms are reproduced well in the case of 14k and approximately in the case of 8k; however, a larger sampling interval leads to a larger error of the binaural signal. In particular, in the cases of 1k and 2k, the waveforms are prominently distorted with unwanted waves whose peak amplitude is larger than that of the main wave. These prominent distortions are attributable to the after-effect wave fronts shown in Figs. 3(e) and 3(f).

Figure 7 shows the frequency characteristics of the resulting binaural signal at the left ear. This figure shows that the frequency characteristics are well reproduced in the case of 14k, whereas a larger sampling interval produces more distorted spectral features. In particular, for the cases of 1k and 2k, the spectral features are highly distorted at higher frequencies, thereby resulting in the complete
destruction of localization cues such as spectral peaks and notches, which are reported to be important in sound image localization. In addition to the reproduction errors at frequencies higher than the upper-bound frequency, there are spectral deviations from "Primary" even at the frequencies below it. These deviations may be attributable to the non-uniform distribution of the sampling points, as described in Sect. 3.1, as well as the approximation error in the estimation of the surface element area $S_1$ in Eq. (2) as described in Sect. 3.2.

Figures 8 and 9 show the ITDs and ILDs, which were calculated from the binaural signals at every 45° azimuth. The ITD was calculated as the delay that maximizes the cross-correlation coefficient between the impulse responses at both ears for the entire frequency range. The ILD was calculated for the entire frequency range. Figure 8 shows that for the cases of 1k and 2k, the ITDs in the reproduced field prominently deviate from those in the primary field. On the other hand, it is indicated that the ITDs are roughly parallel to those in the primary field if the upper-bound frequency is greater than or equal to 4kHz. In addition, Fig. 9 shows that the error in the ILD increases for a larger sampling interval; in particular, large ILD errors ranging from 3 to 10 dB are observed if the upper-bound frequency is less than 4 kHz.

5. DISCUSSION

The numerical results suggest that the ITDs and ILDs in the reproduced field are roughly parallel to those in the primary field if the upper-bound frequency is greater than 4 and 8 kHz, respectively. As demonstrated in Fig. 6, the unwanted after-effect waves, which are more prominent for larger sampling intervals, reach the ear within 1 ms after the arrival of the main wave. According to Blauert [7], if the difference in the arrival time between two coherent sounds is smaller than 1 ms, no precedence effect occurs and a sound image is localized at an average or compromise position between two sounds. In the current results, many wave fronts having relatively large amplitudes reach the ear from various directions within 1 ms after the arrival of the main wave. Therefore, considering the above description, it is likely that, for large sampling intervals, the sound image will not be localized in the direction from which the main wave arrives. Also, for small sampling intervals, there are unwanted after-effect waves from various directions. However, it is likely that their amplitudes will become smaller with decreasing sampling interval, thereby producing a less prominent perceptual impact.

6. CONCLUSIONS

In this paper, the effects of spatial aliasing in sound field reproduction based on integral equations were numerically investigated. The numerical results revealed that a large sampling interval produces unwanted "after-effect" wave fronts reaching the listener in a very short period, 1 ms in the current simulation, after the arrival of the main wave; a larger sampling interval leads to a greater amplitude of the unwanted wave fronts. The values of ITD and ILD in the entire frequency range are approximately reproduced if the upper-bound frequencies are higher than 4 and 8 kHz, respectively.

ACKNOWLEDGMENT

This work was partly supported by SCOPE from MIC, Japan (No. 131202005) and a Grant-in-Aid for Scientific Research (B) from MEXT, Japan (No. 16H02857).

REFERENCES

[1] M. Camras, “Approach to recreating a sound field,” J. Acoust. Soc. Am., 43, 1425–1431 (1968).
[2] A. J. Berkhout, D. de Vries and P. Vogel, “Acoustic control by wave field synthesis,” J. Acoust. Soc. Am., 93, 2764–2778 (1993).
[3] S. Ise, “A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse
systems,” *Acta Acust.* **53**, 706–713 (1997).

[4] M. Otani, H. Watabe, T. Tsuchiya and Y. Iwaya, “Numerical examination of effects of discretization spacing on accuracy of sound field reproduction,” *Acoust. Sci. & Tech.*, **36**, 362–365 (2015).

[5] T. Kimura and K. Kakehi, “Effects of microphone and loudspeaker directivity on accuracy of synthesized wave fronts in sound field reproduction with wave field synthesis,” *Trans. Virtual Reality Soc. Jpn.*, **12**, 191–198 (2007).

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

[7] J. Blauert, *Spatial Hearing*, rev. ed. (MIT Press, Cambridge, 1997).