MULTI-ZONE REPRODUCTION OF A SHARED SOUND FIELD INCLUDING DISTANCE EFFECTS

Taku Shimizu
Tohoku University, Graduate School of Information Sciences and Research Institute of Electrical Communication, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan
Research Fellow of Japan Society for the Promotion of Science, Japan
email: shimizu@ais.riec.tohoku.ac.jp

Jorge Trevino, Shuichi Sakamoto and Yôiti Suzuki
Tohoku University, Graduate School of Information Sciences and Research Institute of Electrical Communication, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

Tomohiko Ise
Alpine Electronics, Inc., 20-1, Yoshima Industrial Park, Iwaki, Fukushima 970-1192, Japan

Expectations of recent audio technologies include the capability to present 3D spatial sound with high levels of realism. Sound field reproduction methods, which physically re-create the sound pressure field of a target acoustic space, outperform other techniques, such as the multi-channel surround systems, from a theoretical stance. However, their implementation often requires a large loudspeaker array with an impractical number of loudspeakers. The situation worsens if multiple listeners are considered since the sound field must be controlled over a region large enough to cover them all. Multi-zone sound field reproduction attempts to sidestep this problem by controlling the sound pressure inside small and disconnected regions, one for each listener. Traditionally, however, multi-zone sound field reproduction systems attempt to present a different acoustic space for each listener. This can impose unrealistic constraints for some listener positions, which affects reproduction accuracy. A shared sound field reproduction method recently proposed by the authors seeks to overcome these limitations. The proposal defines a multi-zone transfer function and uses a phase correction stage to present a single, underlying sound field at multiple disconnected regions. Its formulation is, however, limited to plane waves which cannot convey distance-related information. The present research extends the shared sound field reproduction approach by generalizing the phase correction stage as a translation operator in the spherical harmonics domain. This new signal processing can present a point source field at multiple listening regions, as verified through several numerical simulations evaluating its performance. These simulations assume the presentation of a 1 kHz point source at two listening regions using 192 loudspeakers arranged on an almost regularly sampled sphere. Simulations show the successful reproduction at the target regions, both reaching a reproduction error level below -6 dB. Furthermore, the observation of parallax effects between regions confirms that the proposal can effectively convey distance-related information.

Keywords: sound field control, multi-zone sound field reproduction, sound field reproduction, spatial sound, spherical acoustics

1. Introduction

Advances have been recently made in audio technology for presenting highly accurate spatial sound to listeners. One proposed approach which physically re-creates the sound pressure field is termed sound field reproduction or sound field synthesis [1, 2, 3]. This technique can present high-definition spatial sound to the listener better than other conventional audio technologies such as the multi-channel surround system, but may require many loudspeakers. Practically, this may demand an impractical number of loudspeakers when multiple listeners use a sound field synthesis system simultaneously because of the need for control over an extended listening region.

On the other hand, an approach to re-create sound fields for each of the listeners, named multi-zone sound field reproduction, has been proposed [4, 5]. Multi-zone sound field reproduction seeks to synthesize the individual spatial sound around each listener; this is called *personal sound zone* [6]. Personal sound zones provide significant benefits in that the system presents a different spatial sound for each listener. However, this has several limitations. One of the limitations is that it causes interference when listening areas are close to each other. To overcome the limitations, we have previously proposed a multi-zone sound field reproduction method named *shared sound field synthesis* [7]. This proposed system attempts to re-create multiple portions of a single, underlying sound field in multiple disconnected regions. It focuses on sharing spatial sound information of a common underlying sound field, but at different positions. This proposal was, however, limited to synthesize plane wave sound fields, resulting in the absence of distance effects between sound sources and listeners. To deal with this, we proposed an advanced shared sound field synthesis method which is capable of synthesizing a point source including distance effects [8]. In the present paper, we report evaluation results of this new shared sound field method obtained by numerical simulations to show that the proposed method can provide the distance effects in terms of auditory parallax at the target regions.

2. Shared sound field synthesis including distance effects

In this paper, we assume the coordinate system shown in Fig. 1.

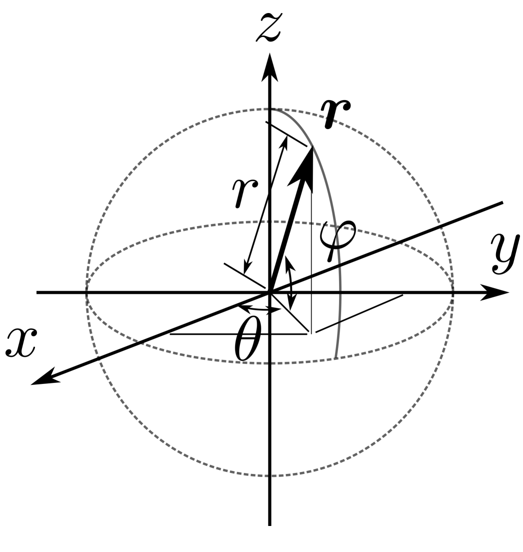


Figure 1: Spherical coordinate system.

Shared sound field synthesis is similar to the reproduction of High-Order Ambisonics (HOA) [2] because both methods use spherical harmonic analysis to decompose the sound field. Here, a sound pressure field p can be decomposed as the following series, where Y_n^m are the spherical harmonic functions and j_n represents the spherical Bessel functions for order n and degree m,

$$p(r, \theta, \varphi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} [B]_n^m(k) j_n(kr) Y_n^m(\theta, \varphi). \tag{1}$$

The set of coefficients $[B]_n^m$ fully characterizes the sound field p. However, the series of Eq. 1 must be truncated at a maximum order N. To synthesize the sound field, shared sound field synthesis uses the mode matching approach [2], expressed as follows,

$$\sum_{l=1}^{L} [C_l]_n^m(k)\, w_l = [B]_n^m(k). \tag{2}$$

Here, w_l denotes the l-th loudspeaker weight $l \in [1, L]$, $k = 2\pi f/c$ represents the wavenumber of a frequency f sound source; the sound speed $c = 343$ m/s. $[C_l]_n^m(k)$ stands for one harmonic component of the transfer function between loudspeaker l located at $\boldsymbol{r}_l = (r_l, \theta_l, \varphi_l)$ and the center of the loudspeaker array. This is expressed by a spherical Hankel function corresponding to an outward spherical wave $h_n^{(\mathrm{out})}$ and the spherical harmonic function Y_n^m as follows:

$$[C_l]_n^m(k) = ikh_n^{(\mathrm{out})}(kr_l)Y_n^m(\theta_l, \varphi_l)^*. \tag{3}$$

Shared sound field synthesis seeks to synthesize sound fields by introducing a translation operator and a multi-zone transfer function [7, 8].

2.1 Translation operator

The translation operator aims to obtain the spherical harmonic coefficients for each of the listening areas from the original one. This operator considers that the observation point for the calculation of the spherical harmonic coefficients is shifted from the center of the loudspeaker array to each listening zone. Assuming the coefficients can be shifted one-by-one, this defines the following formula:

$$[B_z]_n^m(k) = [T_z]_n^m(k)[B_z]_n^m, \tag{4}$$

where the spherical harmonic coefficients of the original source are denoted by $[B]_n^m$ and those of the z-th listening area are written as $[B_z]_n^m$. Here, $[T_z]_n^m(k)$ is an element of the translation operator taking the expansion coefficients at the center of the loudspeaker array to the listening zone z. In this paper, we assume the desired sound field to be that of a point source, and thus the translation operator is expressed as following formula [8]:

$$[T_z]_n^m(k) = \frac{h_n^{(\mathrm{out})}(kr_z^{\mathrm{s}})Y_n^m(\theta_z^{\mathrm{s}}, \varphi_z^{\mathrm{s}})^*}{h_n^{(\mathrm{out})}(kr^{\mathrm{s}})Y_n^m(\theta^{\mathrm{s}}, \varphi^{\mathrm{s}})^*}. \tag{5}$$

Here, the vectors $\boldsymbol{r}_{\mathrm{s}} = (r^{\mathrm{s}}, \theta^{\mathrm{s}}, \varphi^{\mathrm{s}})$ and $\boldsymbol{r}_z^{\mathrm{s}} = (r_z^{\mathrm{s}}, \theta_z^{\mathrm{s}}, \varphi_z^{\mathrm{s}})$ represent the position of the point source and the center of each the listening area, respectively. All coordinates are measured from the center of the loudspeaker array.

2.2 Multi-zone transfer function

The mode matching approach of Eq. 2 aims to re-create a sound field inside a single target area. However, the shared sound field synthesis requires simultaneous control of multiple regions. To achieve this, the shared sound field synthesis proposal uses a multi-zone transfer function obtained by modifying the transfer function of Eq. 2 as follows:

$$[C_l]_n^m(k) \to [C_{l\to z}]_n^m(k). \tag{6}$$

Here, the coefficient $[C_{l\to z}]_n^m(k)$ denotes the transfer function between the loudspeaker l and the listening zone $z \in [1, Z]$ using the following formula:

$$[C_{l\to z}]_n^m(k) = ikh_n^{(\mathrm{out})}(kr_{l\to z})Y_n^m(\theta_{l\to z}, \varphi_{l\to z}), \tag{7}$$

where $\boldsymbol{r}_{l\to z} = (r_{l\to z}, \theta_{l\to z}, \varphi_{l\to z})$ stands for the position of l-th loudspeaker as seen from the z-th listening zone, given here in spherical coordinates.

2.3 Matrix form

Introducing Eq. 5 and Eq. 6 in Eq. 2 yields

$$\sum_{l=1}^{L}\left[[C_{l\to z}]_n^m(k)\right]w_l(k) = [T_z]_n^m(k)[B]_n^m(k). \tag{8}$$

Due to its construction, operator T_z is diagonal, leading to the following formula,

$$\sum_{l=1}^{L}\left[[T_z^{-1}]_n^m(k)[C_{l\to z}]_n^m(k)\right]w_l(k) = [B]_n^m(k). \tag{9}$$

$[T_z^{-1}]_n^m(k)$ stands for an element of the inverse operator of Eq. 5, which is defined by [8]

$$[T_z^{-1}]_n^m(k)[T_z]_n^m(k) = 1, \tag{10}$$

where

$$[T_z^{-1}]_n^m(k) = ([T_z]_n^m(k))^* = \frac{h_n^{(\mathrm{in})}(kr_z^{\mathrm{s}})Y_n^m(\theta_z^{\mathrm{s}},\varphi_z^{\mathrm{s}})}{h_n^{(\mathrm{in})}(kr^{\mathrm{s}})Y_n^m(\theta^{\mathrm{s}},\varphi^{\mathrm{s}})}. \tag{11}$$

Here, $h_n^{(\mathrm{in})}$ represents the spherical Hankel function corresponding to an inward spherical wave of order n and the superscript $*$ denotes complex conjugation. Summarizing the shared sound field synthesis including distance effects for Z zones, is given as

$$\sum_{l=1}^{L}\left[\frac{1}{Z}\sum_{z=1}^{Z}[T_z^{-1}]_n^m(k)[C_{l\to z}]_n^m(k)\right]w_l(k) = [B]_n^m(k). \tag{12}$$

This formula can be expressed in matrix form as follows:

$$\mathrm{C}\boldsymbol{w} = \boldsymbol{B}, \tag{13}$$

where C denotes the $(N+1)^2 \times L$ matrix of the multi-zone transfer functions, $\boldsymbol{w}$ represents the L column vector of loudspeaker signal weights and $\boldsymbol{B}$ is the $(N+1)^2$ column vector of spherical harmonics coefficients. Calculating the Moore-Penrose pseudo-inverse matrix $\mathrm{C}^\dagger$ of the multi-zone transfer function matrix C, we can obtain the following formula for the loudspeaker signal weights to synthesize the sound fields inside multiple regions:

$$\boldsymbol{w} = \mathrm{C}^\dagger\boldsymbol{B}. \tag{14}$$

3. Evaluation

We consider the synthesis of a point source field inside two zones to evaluate the accuracy of the proposed method through numerical simulations. This evaluation assumes a reproduction system which consists of 192 loudspeakers distributed on the surface of a sphere with radius of 2.5 m, as shown in Fig. 2. The reproduction zones are located 0.5 m to the left and right of the center of the loudspeaker array. The spherical harmonic coefficients $[B]_n^m$ are calculated up to a maximum order of 4. Moreover, we define the error between ideal field $p^{(\mathrm{i})}$ and the synthesized field $p^{(\mathrm{r})}$ as follows:

$$E(\boldsymbol{r}) = \max_t\left\{10\log_{10}\left[\left\{p^{(\mathrm{r})}(\boldsymbol{r},t) - p^{(\mathrm{i})}(\boldsymbol{r},t)\right\}|\boldsymbol{r}-\boldsymbol{r}_{\mathrm{s}}|\right]^2\right\}\ [\mathrm{dB}] \\ \left(t = t_0 + \frac{\alpha}{fM}, \alpha = 0,1,\ldots,M-1\right). \tag{15}$$

Here, t is the selected time for simulations, t_0 represents the time it takes for the sound waves radiated by the loudspeakers to cover the entire evaluation zone, and M denotes the number of evaluated instants. Multiplication by $|\boldsymbol{r} - \boldsymbol{r}_\mathrm{s}|$ in Eq. 15 is done to avoid the apparent enlargement of the error due to the attenuation of the wave front as the distance to the sound source increases. In this paper, we assume the number of evaluated instants M to be 5 and the listening zone to be defined as the region where the error, calculated by Eq. 15, does not exceed -6 dB.

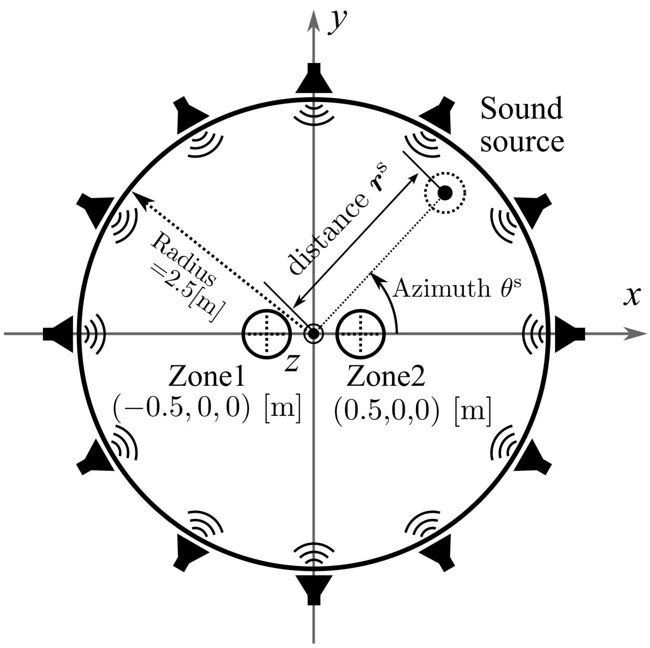


Figure 2: 192-channel spherical loudspeaker array used in numerical simulations. The loudspeakers are assumed to behave as ideal monopoles (point sources) and are distributed on the surface of a sphere with a radius of 2.5 m.

3.1 Simulation

Figure 3 shows the results of reproducing a 1 kHz point source located 2.0 m away from the center of the loudspeaker array. The synthesized sound pressure field and the target field are shown in panel 3(a) and panel 3(b), respectively. Panel 3(c) shows the error calculated by Eq. 15. In these results, the radii of the listening zones are approximately 20 cm and 21 cm.

Next, the simulation considers the effects related to the position of the target point source. Figure 4 shows the radii of the listening zones in relation to the azimuth of the source. The results show that the listening zones are of roughly the same size and independent of the direction of incidence. However, the zones significantly shrink when the zones and source are aligned.

The accuracy in relation to the distance between the center of the loudspeaker array and the target source is shown in Fig. 5. Both listening zones maintain the same size, irrespective of the distance. However, reproduction accuracy decreases rapidly if the translation distance is greater than the distance between the target source and the origin. This implies that the assumptions leading to the proposed translation operator only hold outside a region containing all listening zones.

Finally, we evaluated the performance at multiple frequencies. Figure 6 shows the radii of the listening zones in relation to the target source frequency. The proposed method and HOA achieve a similar performance when a single listening zone is synthesized. The radius of validity R for HOA reproduction is approximated by the rule of thumb

$$R \approx \frac{N}{k}. \tag{16}$$

Comparing Eq. 16 and the results of Fig. 6, it can be confirmed that the proposed method can synthesize sound fields as accurately as single-zone HOA.

4. Conclusion

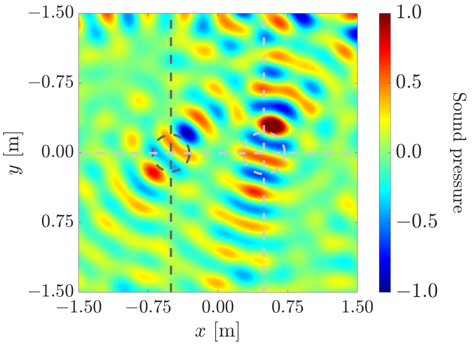
In this paper, we have presented the evaluation of a shared sound field synthesis method, including distance effects, through numerical simulations. These results demonstrate that the direction of incidence does not significantly affect reproduction accuracy, except when the listening zones and target source are aligned. The separation between the sound source and the listening zones also has a negligible impact on sound field reproduction accuracy. Moreover, it was confirmed that the proposed method can synthesize sound fields inside two zones as accurately as single zone reproduction using High-Order Ambisonics.

Acknowledgement

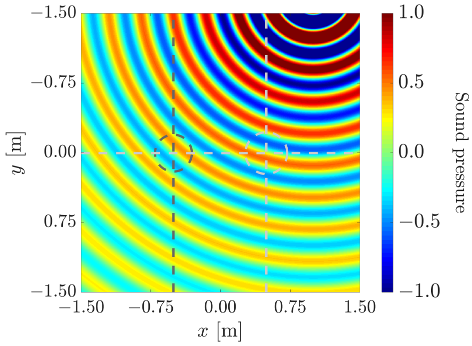
This work was supported by JSPS Grant-in-Aid for JSPS Research Fellow No. 16J07287, JSPS Grant-in-Aid for Scientific Research(A) No. 16H01736, and Aoba Foundation For the Promotion of Engineering.

References

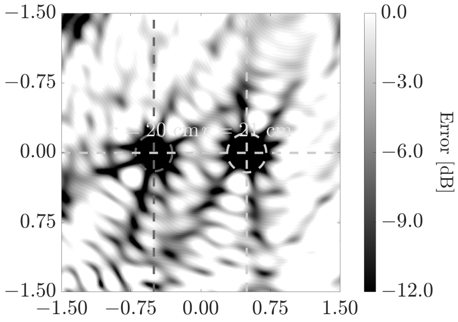
1. Berkhout, A. A holographic approach to acoustic control, *Journal of the Audio Engineering Society*, **36**, 977–995, (1988).
2. Poletti, M. A. Three-Dimensional Surround Sound Systems Based on Spherical Harmonics, *Journal of the Audio Engineering Society*, **53** (11), 1004–1025, (2005).
3. Ise, S. A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems, *Acta Acustica united with Acustica*, **85** (1), 78–87.
4. Poletti, M. An Investigation of 2-D Multizone Surround Sound Systems, *Audio Engineering Society Convention 125*, Oct, pp. 1–9, (2008).
5. Wu, Y. J. and Abhayapala, T. D. Spatial Multizone Soundfield Reproduction: Theory and Design, *IEEE Transactions on Audio, Speech, and Language Processing*, **19** (6), 1711–1720, (2011).
6. Betlehem, T., Zhang, W., Poletti, M. A. and Abhayapala, T. D. Personal Sound Zones: Delivering interface-free audio to multiple listeners, *IEEE Signal Processing Magazine*, **32** (2), 81–91, (2015).
7. Shimizu, T., Trevino, J., Sakamoto, S. and Suzuki, Y. A multi-zone approach to sound field reproduction based on spherical harmonic analysis, *Acoustical Science and Technology*, **36** (5), 441–444, (2015).
8. Shimizu, T., Trevino, J., Sakamoto, S., Suzuki, Y. and Ise, T. Multi-zone Sound Field Reproduction Method Including Distance Effects, *Proceedings of the* 22^{nd} *International Congress on Acoustics*, Sept, (2016).



(a) Synthesized sound pressure field at $\alpha = 1$



(b) Ideal field at $\alpha = 1$



(c) Synthesized error calclated for all of evaluated instants ($M = 5$)

Figure 3: Results of two zone reproduction of a 1 kHz point source, located 2.0 m away from the center of the loudspeaker array at an azimuth of 60 ° and an elevation of 0 °.

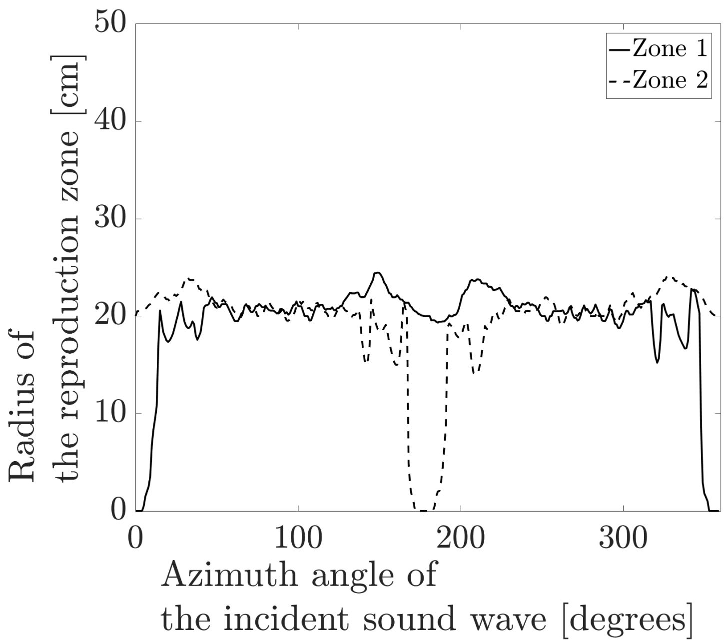


Figure 4: The radii of the listening zones with respect to the direction of the target point source. The azimuth of the source is changed from 0° to 359° every 1°, while remaining at an elevation of 0°.

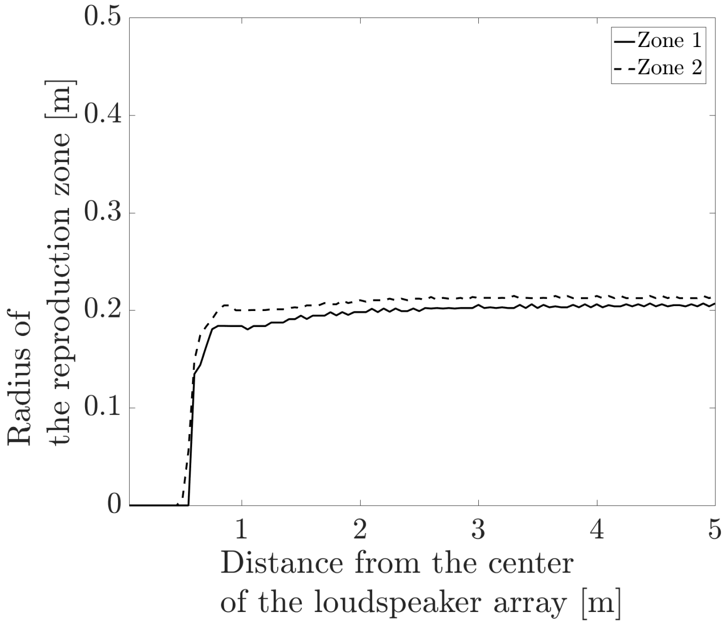


Figure 5: The radii of the listening zones with respect to the distance between the target source and the center of the loudspeaker array. The azimuth and elevation of the source are 60° and 0°, respectively. The distance between the center of the loudspeaker array and the source is changed from 0.05 m to 5 m every 0.05 m.

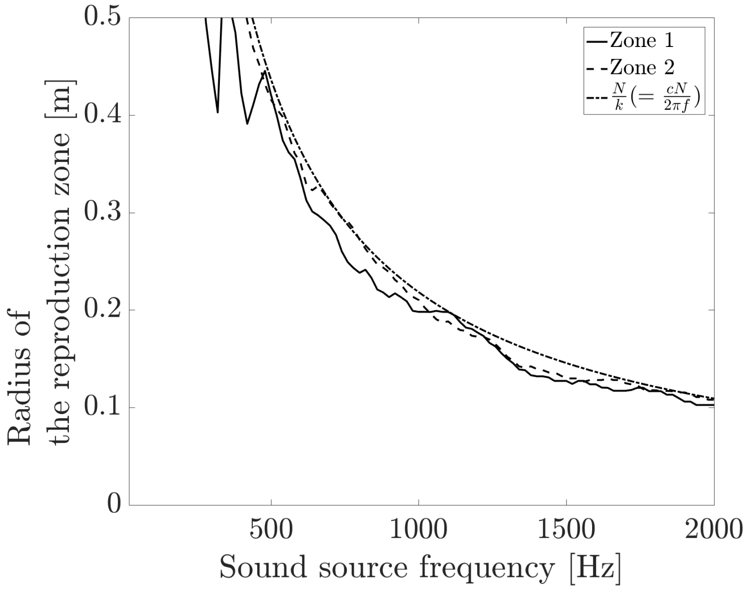


Figure 6: The radii of the listening zones with respect to the target source frequency. The azimuth and elevation of the source are 60° and 0°, respectively. The distance between the center of the loudspeaker array and the source is 2.0 m. The source frequency is changed from 20 Hz to 2000 Hz every 20 Hz.