Audio Engineering Society

Conference Paper

Presented at the Conference on
Headphone Technology
2016 Aug 24–26, Aalborg, Denmark

Numerical evaluation of binaural synthesis from rigid spherical microphone array recordings

César D. Salvador[1], Shuichi Sakamoto[1], Jorge Treviño[1], and Yôiti Suzuki[1]

[1]*Research Institute of Electrical Communication (RIEC) and the Graduate School of Information Sciences (GSIS), Tohoku University, Sendai 980-8577, Japan*

Correspondence should be addressed to César D. Salvador (salvador@ais.riec.tohoku.ac.jp)

ABSTRACT

Binaural systems seek to convey a high-definition listening experience by re-creating the sound pressure at both of the listener's ears. The use of a rigid spherical microphone array (RSMA) allows the capture of sound pressure fields for binaural presentation to multiple listeners. The aim of this paper is to objectively address the question on the required resolution for capturing an individual space. We numerically evaluated how binaural synthesis from RSMA recordings is affected when using different numbers of microphones. Evaluations were based on a human head model. Accurate synthesis of spectral information was possible up to a maximum frequency determined by the number of microphones. Nevertheless, we found that the overall synthesis accuracy could not be indefinitely improved by simply adding more microphones. The limit to the number of microphones beyond which the overall synthesis accuracy did not increase was higher for the interaural spectral information than for the monaural one.

1 Introduction

Binaural technology [1] aims to convey high-definition listening experiences by reproducing the sound pressure signals at both of the listener's ears, namely, *binaural signals*. This allows for the consideration of auditory localization cues that arise from the scattering of sound by the listener's external anatomy [2].

The sound scattering by the listener's external anatomy can be characterized by what are known as the head-related transfer functions (HRTFs) in free field [2]. HRTFs are represented by linear filters relating the position of a sound source and the sound pressure generated by that source at the ears of the listener. To characterize an individual space, HRTF datasets are typically obtained for a spherical array of sound sources [3].

During the last decade, there has been an increasing interest on combining spatial information contained in a HRTF dataset with recordings made with a microphone array [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. In particular, the use of a rigid spherical microphone array (RSMA) provides two main advantages: spatial sound is captured from several directions with uniform resolution, and simultaneous tracking of multiple moving listeners is possible through digital computations. The underlying spherical geometries enable the use of modal representations of HRTF datasets and RSMA recordings, in terms of solutions to the acoustic wave equation at different resolutions (orders). These modal representations enjoy popularity due to their scalability [5, 6, 10, 9, 11, 13, 14, 15, 18, 19].

HRTF datasets can be obtained for high-resolution source distributions using numerical methods [20]. Recently, a perceptual study [21] has reported that low-order HRTF representations might be sufficient to approximate an individual space, whereas an objective study [22] has identified HRTF features that would require high-order representations. Because the required resolution for characterizing an individual space is still an open question, the performance of binaural systems should be evaluated by considering HRTF datasets with the higher resolution that can be achieved.

Existing RSMA recordings, on the other hand, typically contain low-order information only, since high-resolution RSMAs are still hard to construct using actual technology. This has motivated a recent study [11] on adapting the resolutions of HRTF datasets to RSMA recordings by downsampling the HRTF datasets.

Nevertheless, it has also recently been reported in [9] that high-order information is important to synthesize more directionally sharpened and more externalized sounds. Regarding the future of recording technology, RSMAs of hundreds of microphones are not unrealistic. For instance, it has been reported in [23, 24, 12] a setup composed of 252 microphones distributed according to an icosahedral symmetry. Plans to construct higher resolution arrays in the near future also exist.

Bearing these considerations in mind, the present study seek to identify the number of microphones that are necessary to accurately synthesize the spectral information that would be required in binaural localization. We therefore present an extensive numerical evaluation using RSMA recordings and HRTF datasets of different resolutions, up to the amount required to cover all typical audible frequencies objectively. To focus the analysis on the number of microphones, modal representations of RSMA recordings are calculated only. In connection with [11], this is equivalent to resample the RSMA recordings so as to match the resolution of the HRTF dataset. Furthermore, to cope with low-frequency high amplifications typically observed when assuming HRTF datasets characterized for plane-wave sources, we consider the case of point sources, which is more consistent with existing HRTF datasets.

2 Binaural synthesis

In spherical coordinates, a point in space $\vec{r} = (r, \theta, \phi)$ is specified by its radial distance r, azimuth angle $\theta \in [-\pi, \pi]$ and elevation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Angles can be merged into the variable $\Omega = (\theta, \phi)$ in such a way that a point in space is also represented by $\vec{r} = (r, \Omega)$.

For each frequency bin, we define an input vector

$$\mathbf{p} = \begin{bmatrix} p_1 & p_2 & \cdots & p_Q \end{bmatrix}^{\mathsf{T}}. \tag{1}$$

The symbol $^{\mathsf{T}}$ indicates transpose. Each entry p_q of $\mathbf{p}$, where $q = 1, 2, ..., Q$, and Q is the number of microphones, represents a sample in the frequency domain of a sound pressure signal recorded at the microphone position $\vec{r}_q^{\,\mathrm{m}} = (r_\mathrm{m}, \Omega_q^\mathrm{m})$ over a rigid baffle of radius r_m.

We also define a user parameter matrix

$$\mathbf{h} = \begin{bmatrix} h_1^{\mathrm{left}} & h_2^{\mathrm{left}} & \cdots & h_L^{\mathrm{left}} \\ h_1^{\mathrm{right}} & h_2^{\mathrm{right}} & \cdots & h_L^{\mathrm{right}} \end{bmatrix}^{\mathsf{T}} \tag{2}$$

Each entry h_ℓ^{left} or h_ℓ^{right} of $\mathbf{h}$, where $\ell = 1, 2, ..., L$, and L is the number of source positions, represents a sample in frequency of a free-field HRTF for the left or right ear, respectively. Each entry is characterized for a source position $\vec{r}_\ell^{\,\mathrm{v}} = (r_\mathrm{v}, \Omega_\ell^\mathrm{v})$ at a radius r_v. We refer to these positions as the *virtual loudspeaker* positions.

The synthesized *binaural signals* for the left and right ears are organized in the pair

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{b}^{\mathrm{left}} & \hat{b}^{\mathrm{right}} \end{bmatrix}^{\mathsf{T}}. \tag{3}$$

Binaural synthesis can be summarized as the following linear combination of $\mathbf{p}$ and $\mathbf{h}$:

$$\hat{\mathbf{b}} = \mathbf{h}^{\mathsf{T}}\mathbf{A}\mathbf{p}. \tag{4}$$

Here, the combination matrix of size $L \times Q$ is calculated according to the following expression:

$$\mathbf{A} = \mathbf{D}\mathbf{E}^{+}, \tag{5}$$

where $\mathbf{E}^{+}$ and $\mathbf{D}$ denote the encoding and decoding matrices, respectively.

The matrix $\mathbf{E}^{+}$ calculates a representation of $\mathbf{p}$ in terms of harmonic solutions to the acoustic wave equation up to order N. It further compensates for the presence of the spherical baffle, and extrapolates the resulting free-field representation from r_m to r_v. The matrix $\mathbf{E}^{+}$ is obtained by calculating the pseudo-inverse of a matrix $\mathbf{E}$ by using Tikhonov regularization [25]. The entries of $\mathbf{E}$ represent acoustic transfer functions from

an arbitrary position at a distance r_v to each microphone position $\mathbf{r}_q^\mathrm{m}$. The size of $\mathbf{E}$ is $Q \times (N+1)^2$ and its entries, for the assumption of virtual loudspeakers radiating spherical waves, are

$$e_{q,n^2+n+m+1} = \frac{-h_n(kr_\mathrm{v})Y_n^m(\Omega_q^\mathrm{m})}{kr_\mathrm{m}^2 h'_n(kr_\mathrm{m})}. \qquad (6)$$

Here, h_n denotes the spherical Hankel function of the second kind and order n, while Y_n^m denote the complex spherical harmonic functions of order n and degree m, where $n = 0,1,...,N$, and $m = -n,-n+1,...,n$. The functions h_n and Y_n^m are defined in [26], respectively as the radial and angular portions of the solutions to the acoustic wave equation. The symbol $'$ denotes the derivative of a function with respect to its argument. The benefit of performing regularization by including the radial portion is the attenuation of high modal components at low frequencies.

The matrix $\mathbf{D}$ takes the encodings $\mathbf{E}^+\mathbf{p}$ at a radius r_v and decodes them for each virtual loudspeaker directions Ω_ℓ^v. The size of $\mathbf{D}$ is $L \times (N+1)^2$ and has entries

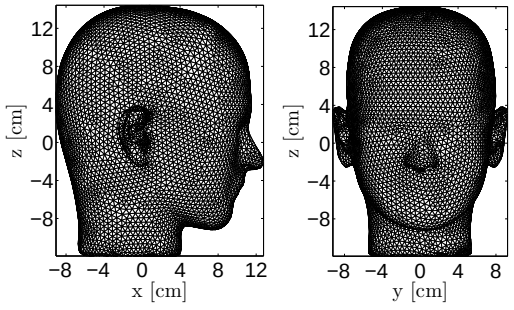
$$d_{\ell,n^2+n+m+1} = \frac{\exp(jkr_\mathrm{v})}{r_\mathrm{v}} Y_n^m(\Omega_\ell^\mathrm{v}). \qquad (7)$$

The fraction represents a free-field transfer function from r_v to the head center. Its purpose is to set the head center as the observation point, in consistency with the definition of free-field HRTFs.

3 Synthesis accuracy

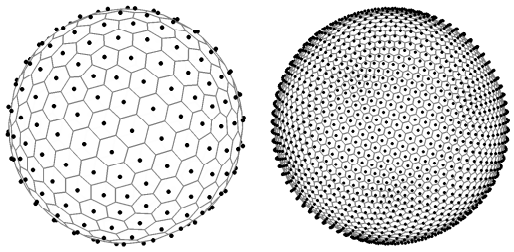
We evaluated the effect of using different numbers of microphones (Q) and virtual loudspeakers (L) on the synthesis accuracy. To emphasize the preservation of spectral information used in human auditory localization, we gave special attention to the synthesis of monaural and interaural spectral information.

Only one example sound source distance (and correspondingly one virtual loudspeaker radius $r_\mathrm{v} = 1.5$ m) was used. Most of the available HRTF datasets are measured at this typical radius, beyond which the HRTFs hardly depend on distance [27]. An exhaustive evaluation at different distances close to the head, while important, is outside the intended scope of this paper.



(a) Lateral view. **(b)** Front view.

Fig. 1: Head model used for numerical experiments.



(a) 252 points. **(b)** 1962 points.

Fig. 2: Examples of icosahedral grids.

3.1 Conditions of the evaluation

Evaluations were based on comparisons of synthesized sets $\hat{B}$ of *binaural transfer functions* that describes a binaural system, and reference sets H of HRTF datasets (target). Synthesized sets $\hat{B} = \{\hat{B}_\mathrm{left}(\Omega_i, f_j), \hat{B}_\mathrm{right}(\Omega_i, f_j)\}$, where $i = 1,2,...,I$, and $j = 1,2,...,J$, were calculated using (4) for the particular case of a number I of point sources placed at a 1.5 m distance, in the directions $\Omega_i = (\theta_i, \phi_i)$. The point sources were radiating non-simultaneously. Each point source was radiating a single sinusoidal signal at a time, and this case was repeated for a number J of single frequencies f_j in the full audible range. On the other hand, reference sets $H = \{H_\mathrm{left}(\Omega_i, f_j), H_\mathrm{right}(\Omega_i, f_j)\}$, were calculated using the boundary element method (BEM) [20] for the head model described in Fig. 1. The mesh grid of this head model consisted of 14,096 points with an average cell length of 5.1 mm, which limited our evaluations to an average frequency of 16.6 kHz.

To calculate $\hat{B}$, microphone signals for the non-simultaneous point sources (input) were first calculated with the algorithm in [28], assuming a rigid sphere of radius $r_\mathrm{m} = 8.5$ cm. Then, HRTF datasets for a virtual loudspeaker array of radius $r_\mathrm{v} = 1.5$ m (user parameter) were calculated using BEM [20] and the head model shown in Fig. 1. The positions of microphones and virtual loudspeakers were decided using spherical grids constructed by subdividing the edges of the icosahedron into equal segments. Examples of such grids are shown in Fig. 2, where dots indicate the positions and lines enclose their Voronoi cells. Although not explicitly mentioned in (6) and (7), encoding and decoding fundamentally lie on numerical integrations on the sphere. The required quadrature weights were thus determined to be proportional to the cell areas.

To select the maximum order N_j required to approximate a frequency f_j, we used the bound proposed in [29]. Maximum orders N_j are determined by setting a constant truncation error within a region of interest enclosed by a given radius. In our simulations, we set a truncation error equal to 10^{-5} within the radius $r_\mathrm{m} = 8.5$ cm of the microphone array. Source distance information is also considered by this rule; we set this variable equal to 1.5 m. Under these conditions, approximations up to an average limit frequency of 16.6 kHz would require an order $N = 43$ and, hence, at least $Q = (43+1)^2 = 1936$ microphones. Conversely, a given number of microphones would limit the spatial resolution up to a maximum frequency.

The regularization parameter required to calculate $\mathbf{E}^+$ was empirically chosen so as to obtain a good compromise between the error and the energy of the source. It was therefore set equal to $\|\mathbf{E}\| \times 10^{-7}$, with the norm of $\mathbf{E}$ equal to its largest singular value.

3.2 Objective measures of accuracy

Sets $\hat{B}$ and H were compared giving special attention to the spectral monaural and interaural spectral information. The monaural local error in decibels is defined by [30]:

$$E_\mathrm{M}(\Omega_i, f_j) = 20\log_{10}\left|\frac{\hat{B}_\mathrm{left}(\Omega_i, f_j)}{H_\mathrm{left}(\Omega_i, f_j)}\right|. \qquad (8)$$

To highlight the capability of synthesizing the main peaks and notches of the HRTFs, which provide important features for auditory localization, the overall gain mismatch was further removed from (8) by subtracting its overall mean value $\bar{E}_\mathrm{M}$.

Because interaural information is important in sound localization, as opposed to monoaural phase [31], interaural measures of accuracy were also considered.

The reference interaural HRTFs are defined as [2]

$$H_\mathrm{interaural}(\Omega_i, f_j) = \frac{H_\mathrm{left}(\Omega_i, f_j)}{H_\mathrm{right}(\Omega_i, f_j)}, \qquad (9)$$

and the synthesized interaural transfer functions as

$$\hat{B}_\mathrm{interaural}(\Omega_i, f_j) = \frac{\hat{B}_\mathrm{left}(\Omega_i, f_j)}{\hat{B}_\mathrm{right}(\Omega_i, f_j)}. \qquad (10)$$

The interaural level differences (ILDs) corresponding to the reference and synthesized transfer functions are defined by the magnitude in decibels of (9) and (10), respectively. We calculated the ILD local error correspondingly in decibels by

$$E_\mathrm{ILD}(\Omega_i, f_j) = 20\log_{10}\left|\frac{\hat{B}_\mathrm{interaural}(\Omega_i, f_j)}{H_\mathrm{interaural}(\Omega_i, f_j)}\right|. \qquad (11)$$

On the other hand, the interaural phase differences (IPDs) associated with the reference and synthesized transfer functions correspond to the phase in radians of (9) and (10), respectively. In particular, phase information can be displayed by means of its group delay to highlight spectral information related to peaks and notches with a better resolution [32]. We calculated the interaural group delay (IGD) local error correspondingly in seconds according to

$$E_\mathrm{IGD}(\Omega_i, f_j) = \frac{\Delta\arg\left(\frac{\hat{B}_\mathrm{interaural}(\Omega_i, f_j)}{H_\mathrm{interaural}(\Omega_i, f_j)}\right)}{2\pi\Delta f_j}, \qquad (12)$$

where arg denotes the unwrapped phase and Δ is the finite difference operator along the discrete variable f_j.

We examined the overall accuracy of our method based on the root mean square (RMS) values of the errors defined in (8), (11) and (12). Norms of accuracy equivalent to the RMS value have been evaluated through listening tests in [30], where the suitability of these norms for predicting audible differences between measured and synthesized HRTFs was verified. We calculated

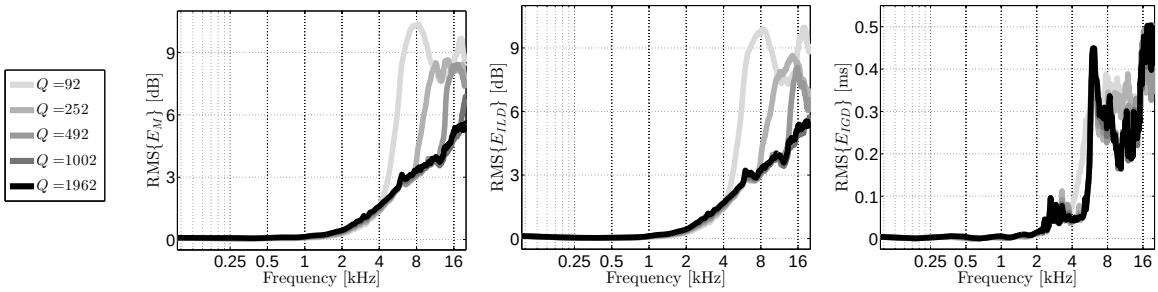


Fig. 3: Overall accuracy on the sphere calculated using (13), for synthesis with $L = 1962$ virtual loudspeakers and different numbers of microphones (Q).

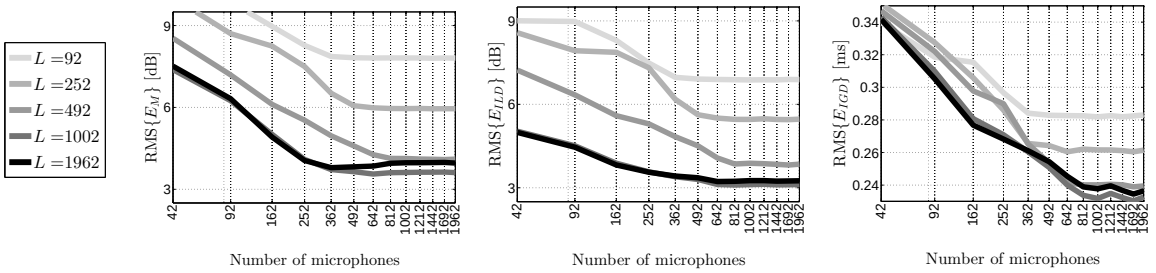


Fig. 4: Overall accuracy on the sphere calculated using (14) up to 16.6 kHz, for synthesis with different numbers of virtual loudspeakers (L).

the RMS value along directions based on the following expression:

$$\underset{\Omega_i}{\mathrm{RMS}}\{E\}(f_j) = \left(\sum_{i=1}^{I} E(\Omega_i, f_j)^2 w_i\right)^{\frac{1}{2}}. \qquad (13)$$

Here, E can be one of the errors E_{M} in (8), E_{ILD} in (11), or E_{IGD} in (12), and w_i are normalized quadrature weights (area of Voronoi cells) for numerical integration over all of the sound source directions on a sphere with a radius of 1.5 m. Similarly, we extended the calculation of the RMS value to cover all directions and frequencies according to:

$$\underset{\Omega_i, f_j}{\mathrm{RMS}}\{E\} = \left(\frac{1}{J}\sum_{i,j=1}^{I,J} E(\Omega_i, f_j)^2 w_i\right)^{\frac{1}{2}}. \qquad (14)$$

3.3 Synthesis on the sphere

We considered $I = 5762$ sound sources almost uniformly distributed on a sphere with a radius of 1.5 m and frequency bins in the full audible range for a sampling frequency of 48 kHz. The results based on (13) and (14) are displayed in Figs. 3 and 4, respectively.

In Fig. 3, it can be observed that, when a number of virtual loudspeakers sufficient to cover the audible frequency range was used, excellent monaural and interaural overall accuracies were obtained up to 2 kHz, even with a limited number of microphones. Over 2 kHz, accuracies of monaural levels and interaural level differences (left and middle panels) gradually decreased with increasing frequency and a decreasing number of microphones. Nevertheless, increasing the number of microphones beyond 1002 did not yield a significant improvement in these overall accuracies. Regarding the interaural group delay (right panel), good performance was also obtained at low frequencies, precisely where this spectral information is known to be important.

In Fig. 4, it can be observed that both monaural and interaural accuracies were also degraded when the number of virtual loudspeakers decreased below 1002. However, increasing this number over 1002 did not improve the overall accuracies. On the other hand, overall

accuracies were significantly improved by increasing the number of microphones up to a certain limit, after which adding more microphones did not lead to a decrease in the RMS errors. For a number of virtual loudspeakers greater than 1002, the left panel shows that increasing the number of microphones beyond 362 did not improve the overall accuracy of the monaural information. For the same condition, the middle panel shows that the overall accuracy in ILD did not benefit from additional microphones beyond 642. Furthermore, for more than 1002 virtual loudspeakers, the right panel shows that increasing the number of microphones beyond around 1002 did not improve the overall IGD accuracy. The limits were different depending on the type of spectral information under consideration.

3.4 Evaluation of white noise gain

Binaural synthesis as described in (4) can also be regarded as a filter-and-sum beamformer with weight vector $\mathbf{h}^{\mathrm{T}}\mathbf{A}$. The norm squared of the weight vectors defines the white noise gain:

$$\mathrm{WNG} = \|\mathbf{h}^{\mathrm{T}}\mathbf{A}\|^2, \tag{15}$$

which is an estimate of the output power due to spatially uncorrelated, unit variance white noise at the sensors [33]. If WNG is large, it is expected a poor signal-to-noise ratio at the output of the beamformer due to white noise contributions. The inverse of the white noise gain, WNG^{-1}, is used as an estimator for robustness to noise.

Equating (6) and (7), and using the analytic Tikhonov regularized solution [25], it can be shown that the entries of $\mathbf{A} = [a_{\ell q}]$ in (4) are given by

$$a_{\ell q} = \frac{1}{4\pi(N+1)^2} \cdot \frac{\exp(jkr_{\mathrm{v}})}{r_{\mathrm{v}}} \cdot \sum_{n=0}^{(Q+1)^2} (2n+1) R_n^{\mathrm{reg}}(r_{\mathrm{m}}, r_{\mathrm{v}}, k) P_n(\cos\Theta_{\ell q}). \tag{16}$$

The angular part of the sum in (16) is defined by the Legendre polynomial P_n of order n evaluated at the cosine of the angle $\Theta_{\ell q}$ between $\vec{r}_\ell^{\,\mathrm{v}}$ and $\vec{r}_q^{\,\mathrm{m}}$. The radial part is defined by the regularized radial filter

$$R_n^{\mathrm{reg}} = \frac{R_n}{1+\lambda^2|R_n|^2}, \quad R_n = -\frac{kr_{\mathrm{m}}^2 h_n'(kr_{\mathrm{m}})}{h_n(kr_{\mathrm{v}})}, \tag{17}$$

where λ is the regularization parameter.

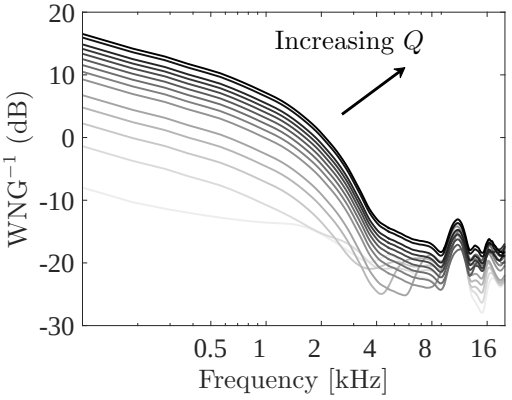


Fig. 5: Inverse of white noise gain for $L = 1962$ virtual loudspeakers and different numbers of microphones (Q =12, 42, 92, 162, 252, 362, 492, 642, 812, 1002, 1212, 1442, 1692, 1962.).

Figure 5 shows the results obtained when calculating WNG^{-1} using the model in (16) with $\lambda = 1 \times 10^{-3}$. It can be observed that increasing the number of microphones improved the signal-to-noise ratio at lower and middle frequencies. However, when the number of microphones increased, its contribution to robustness only increased slightly. This was particularly true at higher frequencies.

4 Conclusion

Based on the simulation of a head model valid up to 16.6 kHz, the effects of using different numbers of microphones and virtual loudspeakers on binaural synthesis were evaluated. The placement of microphones and virtual loudspeakers was determined by following a sampling of the sphere based on the geometry of an icosahedron. Accuracy was evaluated for dense sets of sound source directions. Overall errors for monaural and interaural spectral information were calculated.

In general terms, bounded and smooth synthesis of monaural and interaural spectral features was possible by frequency-dependent order limitation and regularization. The performance of binaural synthesis based on spherical microphone arrays was found to depend mainly on the number of microphones; this determines the maximum frequency that can be resolved by the system. Nevertheless, our results showed a limit after which increasing the number of microphones does not lead to an improvement in accuracy. Furthermore,

different limits were found depending on the type of spectral feature under consideration.

When the number of virtual loudspeakers was sufficiently large to cover frequencies up to 16.6 kHz, we found that the number of microphones required to improve the overall synthesis accuracy of the interaural level difference was higher than the number required to improve the overall synthesis accuracy of the monaural level. Furthermore, the number of microphones required to accurately synthesize the interaural group delay was greater than the number required by the interaural level difference.

Further considerations regarding the synthesis of individual HRTFs, as well as perceptual evaluations by means of detectability of differences, and localization tests along azimuth and elevation angles, could provide more insight into the validity of the present results.

Acknowledgment

This study was supported by a JSPS Grant-in-Aid for Scientific Research (no. 24240016 and 16H01736) and the Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet." The authors wish to thank Makoto Otani for developing the BEM solver used to calculate HRTF datasets, and Junfeng Li for fruitful discussion.

References

[1] Møller, H., "Fundamentals of binaural technology," *Appl. Acoust.*, 36(3–4), pp. 171–218, 1992.

[2] Blauert, J., *Spatial hearing: The psychophysics of human sound localization*, MIT Press, Cambridge, MA, USA; London, England., revised edition, 1997.

[3] Watanabe, K., Iwaya, Y., Suzuki, Y., Takane, S., and Sato, S., "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoust. Sci. Technol.*, 35(3), pp. 159–165, 2014.

[4] Algazi, V. R., Duda, R. O., and Thompson, D. M., "Motion-tracked binaural sound," *J. Audio Eng. Soc.*, 52(11), pp. 1142–1156, 2004.

[5] Duraiswami, R., Zotkin, D. N., Li, Z., Grassi, E., Gumerov, N. A., and Davis, L. S., "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *AES 119*, New York, USA, 2005.

[6] Song, W., Ellermeier, W., and Hald, J., "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *J. Acoust. Soc. Am.*, 123(2), pp. 910–924, 2008.

[7] Laitinen, M. V. and Pulkki, V., "Binaural reproduction for Directional Audio Coding," in *Proc. IEEE WASPAA*, pp. 337–340, Espoo, Finland, 2009.

[8] Rasumow, E., Blau, M., Doclo, S., Hansen, M., Van de Par, S., Püschel, D., and Mellert, V., "Least squares versus non-linear cost functions for a vitual artificial head," *Proc. Mtgs. Acoust.*, 19(1), 2013.

[9] Avni, A., Ahrens, J., Geier, M., Spors, S., Wierstorf, H., and Rafaely, B., "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *J. Acoust. Soc. Am.*, 133(5), pp. 2711–2721, 2013.

[10] Salvador, C. D., Sakamoto, S., Treviño, J., Li, J., Yan, Y., and Suzuki, Y., "Accuracy of head-related transfer functions synthesized with spherical microphone arrays," *Proc. Mtgs. Acoust.*, 19(1), 2013.

[11] Bernschütz, B., Giner, A. V., Pörschmann, C., and Arend, J., "Binaural Reproduction of Plane Waves With Reduced Modal Order," *Acta Acust. United Ac.*, 100(5), pp. 972–983, 2014.

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

[13] Sheaffer, J., van Walstijn, M., Rafaely, B., and Kowalczyk, K., "Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 23(12), pp. 2125–2135, 2015.

[14] Shabtai, N. and Rafaely, B., "Generalized spherical array beamforming for binaural speech reproduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 22(1), pp. 238–247, 2014.

[15] Shabtai, N. R., "Optimization of the directivity in binaural sound reproduction beamforming," *J. Acoust. Soc. Am.*, 138(5), pp. 3118–3128, 2015.

[16] Delikaris-Manias, S., Vilkamo, J., and Pulkki, V., "Parametric binaural rendering utilizing compact microphone arrays," in *Proc. IEEE ICASSP*, pp. 629–633, 2015.

[17] Rasumow, E., Hansen, M., van de Par, S., Puschel, D., Mellert, V., Doclo, S., and Blau, M., "Regularization approaches for synthesizing HRTF directivity patterns," *IEEE/ACM Trans. Audio, Speech, Language Process.*, PP(99), pp. 215 – 225, 2015.

[18] Alon, D. L., Sheaffer, J., and Rafaely, B., "Plane-Wave Decomposition with Aliasing Cancellation for Binaural Sound Reproduction," in *Audio Engineering Society Convention 139*, 2015.

[19] Alon, D. and Rafaely, B., "Beamforming with Optimal Aliasing Cancellation in Spherical Microphone Arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 24(1), pp. 196–210, 2016.

[20] Otani, M. and Ise, S., "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, 119(5), pp. 2589–2598, 2006.

[21] Romigh, G., Brungart, D., Stern, R., and Simpson, B., "Efficient Real Spherical Harmonic Representation of Head-Related Transfer Functions," *IEEE J. Sel. Topics Signal Process.*, 9(5), pp. 921–930, 2015, ISSN 1932-4553.

[22] Bates, A., Khalid, Z., and Kennedy, R., "Novel sampling scheme on the sphere for head-related transfer function measurements," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 23(6), pp. 1068–1081, 2015.

[23] Sakamoto, S., Hongo, S., Kadoi, R., and Suzuki, Y., "SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," in *Proc. 2nd Int. Symp. Universal Comm.*, pp. 429–434, 2008.

[24] Sakamoto, S., Hongo, S., and Suzuki, Y., "3D sound-space sensing method based on numerous symmetrically arranged microphones," *IEICE Trans. Fundamentals*, E97-A(9), pp. 1893–1901, 2014.

[25] Morozov, V. A., *Regularization Methods for Ill-posed problems*, CRC Press, 1993.

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

[27] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, 106(3), pp. 1465–1479, 1999.

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

[29] Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N., "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *J. Acoust. Soc. Am.*, 127(1), pp. 370–386, 2010.

[30] Lee, K.-S. and Lee, S.-P., "A relevant distance criterion for interpolation of head-related transfer functions," *IEEE Trans. Audio, Speech, Language Process.*, 19(6), pp. 1780–1790, 2011.

[31] Kulkarni, A., Isabelle, S. K., and Colburn, H. S., "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.*, 105(5), pp. 2821–2840, 1999.

[32] Raykar, V. C., Duraiswami, R., and Yegnanarayana, B., "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, 118(1), pp. 364–374, 2005.

[33] Van Veen, B. and Buckley, K., "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP*, 5(2), pp. 4–24, 1988, this Magazine ceased production in 1990. The current retitled publication is IEEE Signal Processing Magazine.