Audio Engineering Society

Conference Paper

Presented at the Conference on
Audio for Virtual and Augmented Reality
2018 August 20 – 22, Redmond, WA, USA

Enhancing binaural reconstruction from rigid circular microphone array recordings by using virtual microphones

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

[1]*Research Institute of Electrical Communication (RIEC) and 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

Spatially accurate binaural reconstruction from rigid circular arrays requires a large number of microphones. However, physically adding microphones to available arrays is not always feasible. In environments such as conference rooms or concert halls, prior knowledge regarding source positions allows for the prediction of pressure signals at positions without microphones. Prediction is performed by relying on a physical model for the acoustically rigid sphere. Recently, we used this model to formulate a surface pressure interpolation method for virtual microphone generation. In this study, we use virtual microphones to enhance the high-frequency spatial accuracy of binaural reconstruction. Numerical experiments in anechoic and reverberant conditions demonstrate that adding virtual microphones extends the frequency range of operation and attenuates the time-domain artifacts.

1 Introduction

Virtual reality systems are key components to create high-definition multimodal interfaces for telepresence applications such as teleconferencing, live streaming of concerts, and motion simulators [1, 2, 3]. In these applications, an accurate reconstruction of the spatial features of sound enhances the user's sensation of immersion thanks to the synergy of spatial sound information with information for other modes of perception such as vision, touch, and body vibration [4, 1, 2].

Spatial sound systems for the reconstruction of acoustic pressure signals on the eardrums, referred to as binaural signals, are in high demand because of the popularity of spatial audio reproduction devices for personal use, such as headphones [5] and personal sound zone systems [6]. An important characteristic of a binaural system is its spatial accuracy, that is, its ability to resolve sounds along directions. Spatial accuracy has a high impact on the degree of realism and naturalness when attempting to re-create the experience of being immersed in a distinct acoustic environment [7].

Recent binaural systems [8, 9, 10, 11, 12, 13, 14, 15] aim to reconstruct binaural signals by combining head-related transfer functions (HRTFs) and spatial sound recordings obtained with a microphone array mounted on a rigid spherical baffle. The use of HRTFs allows to account for the individual auditory localization cues that arise from the interactions of sound waves with the external anatomical shapes of listeners [16]. Additionally, the use of microphone arrays allows to account for the dynamic auditory cues that correspond to head movements [8, 9, 10, 11, 12, 13, 14, 15].

Obtaining HRTFs for dense sets of positions is becoming possible by interpolation from measurements for a single distance [17, 18]. More recently, HRTFs can also be calculated with techniques based on 3D model acquisition systems and numerical acoustic methods [19, 20]. In contrast, the recording of high-definition spatial sound still represents a challenge because it requires a large number of microphones.

The costs involved in the construction of high spatial resolution arrays have largely confined their use to research purposes [9, 10, 11, 21]. For instance, the 252-channel system in [10, 11] corresponds to the highest spatial resolution reported as a real-time binaural system. On the other hand, the resolutions of commercially available arrays are still in the range of a few tens of microphones.

Various interpolation methods for generating recording signals at positions on the rigid baffle where there are no microphones have been proposed [8, 22]. These methods, however, require distinct panning procedures at different frequency bands [23]. Alternatively, we have been exploring the conditions under which a single method could be used over the full frequency range.

In environments such as conference rooms or concert halls, the positions of sound sources are often confined to a small region within the space in front of the listeners. In these conditions, where sound source positions can be assumed to be known, the pressure generated at any point on the rigid spherical baffle can be estimated using a physical model [24]. This model is valid over the full frequency range. Based on this model, we recently proposed a virtual microphone generation method that leverages prior knowledge regarding source positions [25].

In this study, we apply the virtual microphone generation method in [25] to enhance the high-frequency spatial accuracy of binaural reconstruction. The transition between lower and higher frequencies is determined by the number of available microphones. In this paper, we review the binaural reconstruction algorithm used in [10, 11, 15] and the virtual microphone generation method used in [25]. Subsequently, we present numerical results of binaural reconstruction performed in anechoic and reverberant conditions.

2 Binaural reconstruction

One approach to binaural reconstruction can be regarded as a beamformer [10, 11, 15]. An individual HRTF dataset ($\boldsymbol{h}$) constitutes a specified spatial pattern to be approximated by a set of weighting filters ($\boldsymbol{w}$) that are applied to the microphone recordings ($\boldsymbol{p}$). The weighting filters are created by solving a linear system of equations that approximates the HRTF dataset. The entries in the matrix ($\boldsymbol{C}$) associated with the linear system are acoustic transfer functions from the positions of microphones to the positions used to obtain the HRTF dataset.

Let M be the number of microphones and L the number of HRTF positions. The linear system to be solved for each ear is formulated as:

$$\boldsymbol{C}_{L\times M}\cdot\boldsymbol{w}_{M\times 1}=\boldsymbol{h}_{L\times 1}+\boldsymbol{\varepsilon}_{L\times 1}, \tag{1}$$

where $\boldsymbol{\varepsilon}$ denotes approximation error. The solution to (1) leads to a two-step reconstruction algorithm:

1. Weighting filters: $\boldsymbol{w}=\boldsymbol{C}^{+}\boldsymbol{h}$.
2. Binaural signals: $\boldsymbol{b}=\overline{\boldsymbol{w}^{\top}}\boldsymbol{p}$.

The reconstructed binaural signals $\boldsymbol{b}$ constitute the input to binaural reproduction devices based on headphones or personal sound zone systems. The symbols $\cdot^{+}$, $\cdot^{\top}$, and $\overline{\cdot}$ denote pseudo-inversion, transpose, and complex conjugate, respectively. Truncated singular value decomposition is used in this study to perform pseudo-inversion. The above-described reconstruction procedure results in the processing structure illustrated in Fig. 1.

3 Virtual microphone generation

The method introduced in [25] is summarized in this section. Consider a rigid sphere with radius a and a sound source at a distance $r>a$ measured from the center of the rigid sphere. The total pressure Ψ generated by a source placed at a position $\vec{r}=(r,\theta)$ and measured by an ideal microphone placed at a position $\vec{a}=(a,\theta_a)$ on the rigid spherical surface is defined as [24]:

$$\Psi(\vec{r},\vec{a},k)=\frac{-1}{ka^2}\sum_{n=0}^{\infty}\frac{h_n(kr)}{h_n'(ka)}(2n+1)L_n(\cos\Theta_{\vec{r},\vec{a}}). \tag{2}$$

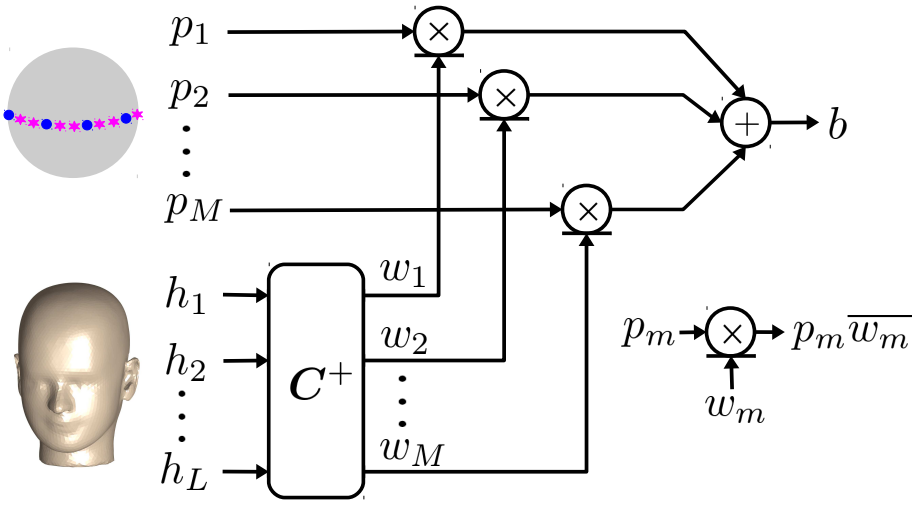


Fig. 1: Binaural reconstruction system as described in [10, 11, 15]. Virtual microphone generation [25] is used to increase the inputs (p_m) to this system. The total number of microphones $M = M_\mathrm{r} + M_\mathrm{v}$ consists of M_r real microphones (blue dots) and M_v virtual microphones (magenta marks).

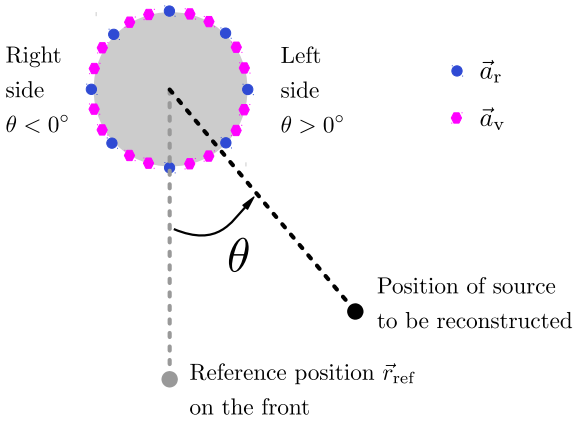


Fig. 2: Geometry for virtual microphone generation (top view). The angle θ is the azimuthal angle on the horizontal plane. The front is at $\theta = 0°$ and the left ear is at $\theta = 90°$. A single reference position $\vec{r}_\mathrm{ref}$ on the front is assumed known.

Here, k denotes the wave number, h_n denotes the spherical Hankel function of order n, and the symbol $'$ indicates a derivative with respect to the argument. Additionally, L_n denotes the Legendre polynomial of order n evaluated at the cosine of the angle $\Theta_{\vec{r},\vec{a}}$ between $\vec{r}$ and $\vec{a}$.

When a reference sound source position $\vec{r}_\mathrm{ref}$ is known, the model in (2) can be used to relate the pressure at two arbitrary points $\vec{a}_\mathrm{r}$ and $\vec{a}_\mathrm{v}$ on the rigid sphere. This defines the surface pressure variation filter as:

$$F_{\mathrm{r}\mapsto\mathrm{v}} = \frac{\Psi_\mathrm{v}}{\Psi_\mathrm{r}}, \tag{3}$$

where

$$\Psi_\mathrm{r} = \Psi(\vec{r}_\mathrm{ref}, \vec{a}_\mathrm{r}, k), \text{ and } \Psi_\mathrm{v} = \Psi(\vec{r}_\mathrm{ref}, \vec{a}_\mathrm{v}, k). \tag{4}$$

The filter $F_{\mathrm{r}\mapsto\mathrm{v}}$ represents the transmission of sound over the rigid spherical surface from $\vec{a}_\mathrm{r}$ to $\vec{a}_\mathrm{v}$.

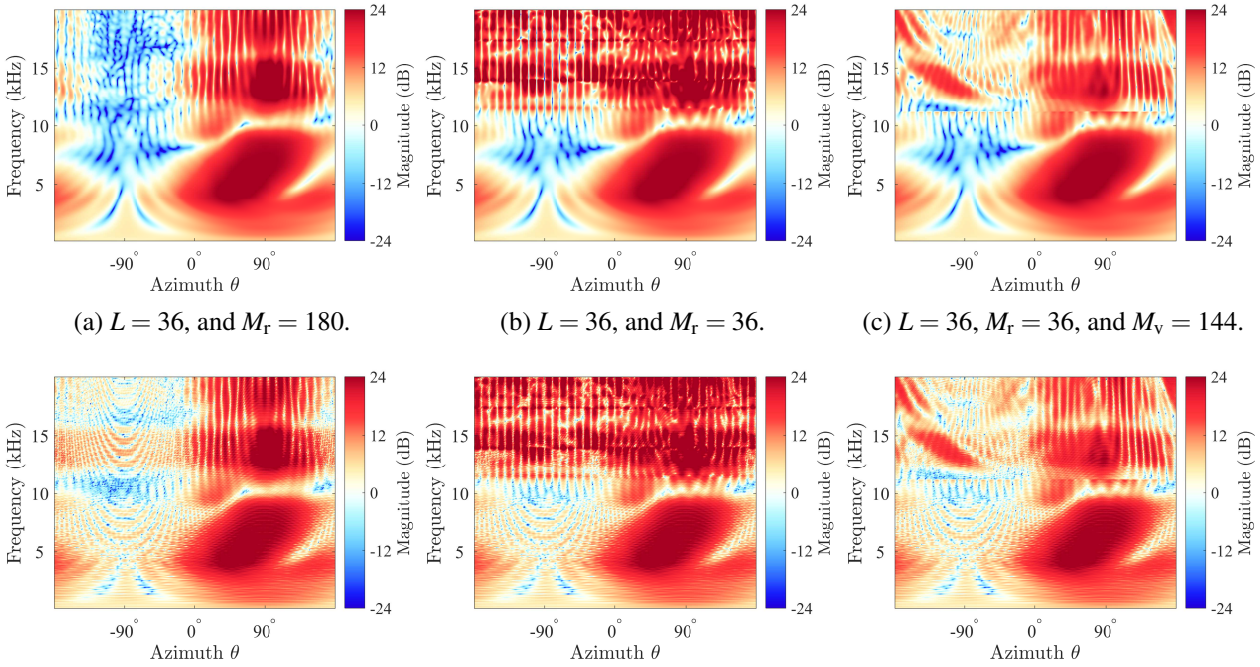
Let p_r denote the signal recorded by a real microphone placed at a position $\vec{a}_\mathrm{r}$. Let p_v denote a virtual microphone signal to be generated at a position $\vec{a}_\mathrm{v}$ where there are no real microphones. The signal p_v is obtained as follows:

$$p_\mathrm{v} = F_{\mathrm{r}\mapsto\mathrm{v}} \times p_\mathrm{r}. \tag{5}$$

Note that the use of $F_{\mathrm{r}\mapsto\mathrm{v}}$ in (3) only requires to specify $\vec{r}_\mathrm{ref}$ and $\vec{a}_\mathrm{v}$. The position $\vec{a}_\mathrm{r}$ from which interpolation is performed can be easily obtained using a nearest neighbor search that consists of two steps. First, the angles between $\vec{a}_\mathrm{v}$ and every $\vec{a}_\mathrm{r}$ are calculated. Next, the real microphone position $\vec{a}_\mathrm{r}$ that creates the smallest angle from $\vec{a}_\mathrm{v}$ is selected.

Given a number M_r of real microphones in the array, its frequency range of operation is delimited by [26, 27]

$$f_\mathrm{max} \approx \frac{M_\mathrm{r} c}{4\pi a}, \tag{6}$$



(a) $L = 36$, and $M_\mathrm{r} = 180$. (b) $L = 36$, and $M_\mathrm{r} = 36$. (c) $L = 36$, $M_\mathrm{r} = 36$, and $M_\mathrm{v} = 144$.

(d) $L = 36$, and $M_\mathrm{r} = 180$. (e) $L = 36$, and $M_\mathrm{r} = 36$. (f) $L = 36$, $M_\mathrm{r} = 36$, and $M_\mathrm{v} = 144$.

Fig. 3: Left-ear transfer functions reconstructed from recordings in anechoic (top) and reverberant (bottom) conditions for $\vec{r}_\mathrm{ref} = (50\ \mathrm{cm}, 0^\circ)$. (a) and (d): *Target*. (b) and (e): *Real*. (c) and (f): *Real+Virtual*.

where c denotes the speed of sound in air. The maximum frequency f_max determines the transition between the lower and higher frequencies. Reconstruction artifacts above f_max are referred to in the literature as spatial aliasing [26, 27].

Virtual microphone generation aims at increasing f_max by using prior knowledge of source positions. Because lower frequencies are sufficiently covered by the available number of real microphones, they need to remain unmodified. Virtual microphones are therefore generated for the higher frequencies only.

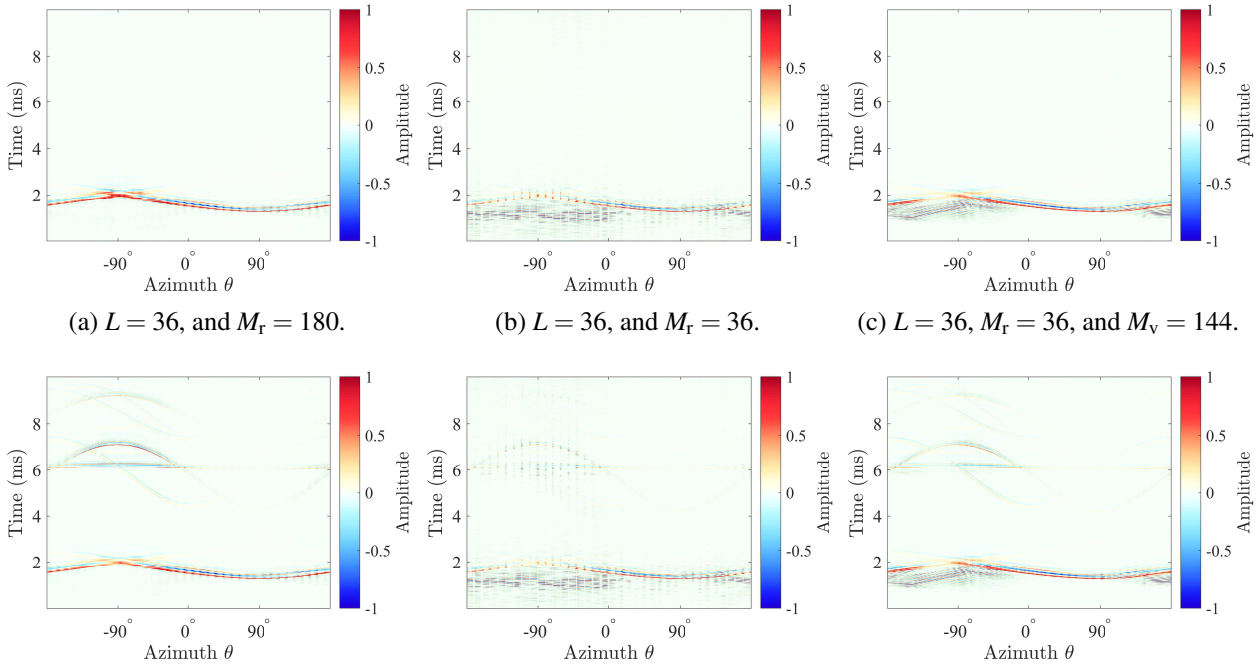
4 Numerical evaluation

In this section, the virtual microphone generation method described in section 3 is used to increase the number of real-microphone recording signals that feed the binaural reconstruction system described in section 2 (see Fig. 1). Virtual microphones are only added at higher frequencies, which is the range that can not be covered by the limited number of real microphones.

The numerical evaluations hereafter consider a rigid spherical baffle of radius $a = 8.5$ cm. The following three conditions regarding the number of microphones were considered:

- *Target*: A number $M_\mathrm{r} = 180$ of real microphones was selected as the target condition.
- *Real*: A number $M_\mathrm{r} = 36$ of real microphones was used to exemplify a practical situation.
- *Real+Virtual*: At frequencies above $f_\mathrm{max} \approx 11$ Khz, a number $M_\mathrm{v} = 144$ of virtual microphones was added to the *Real* case $M_\mathrm{r} = 36$ so as to obtain a total number of microphones $M = M_\mathrm{r} + M_\mathrm{v} = 180$.

All evaluations used a number $L = 36$ of HRTF positions at a distance of 50 cm from the center of the head. All microphone and HRTF positions were equiangularly distributed on horizontal circles. The reference position $\vec{r}_\mathrm{ref}$, which is used to calculate $F_{\mathrm{r}\mapsto\mathrm{v}}$, was placed in front of the listener ($\theta = 0^\circ$) at a distance of



(a) $L = 36$, and $M_\mathrm{r} = 180$. (b) $L = 36$, and $M_\mathrm{r} = 36$. (c) $L = 36$, $M_\mathrm{r} = 36$, and $M_\mathrm{v} = 144$.

(d) $L = 36$, and $M_\mathrm{r} = 180$. (e) $L = 36$, and $M_\mathrm{r} = 36$. (f) $L = 36$, $M_\mathrm{r} = 36$, and $M_\mathrm{v} = 144$.

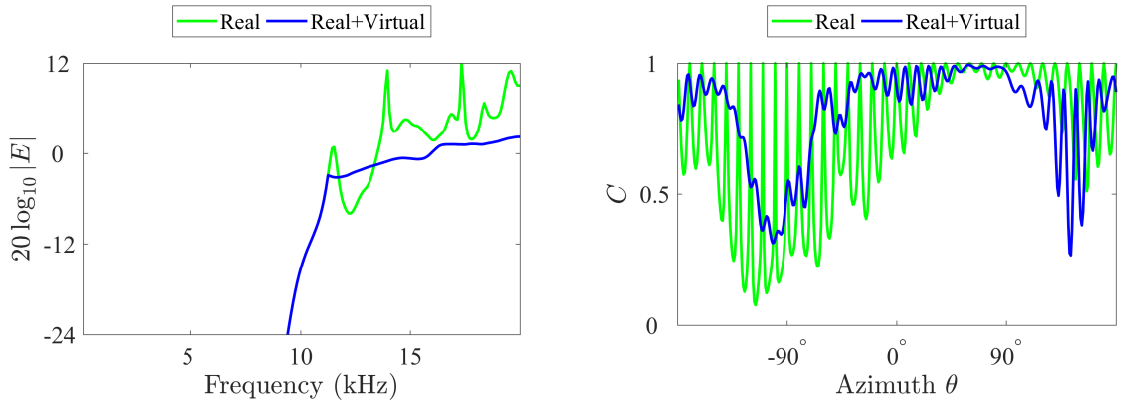
Fig. 4: Left-ear impulse responses reconstructed from recordings in anechoic (top) and reverberant (bottom) conditions for $\vec{r}_\mathrm{ref} = (50\ \mathrm{cm}, 0^\circ)$. (a) and (d): *Target*. (b) and (e): *Real*. (c) and (f): *Real+Virtual*.

50 cm. The sources to be recorded were also placed at a distance of 50 cm (see Fig. 2).

Recordings in reverberant conditions were simulated using the algorithm in [28]. This algorithm calculates the sound pressure on the surface of a rigid sphere that is placed inside a rectangular parallelepiped room. The calculated sound pressure corresponds to the total reverberant field, which includes the scattering from the rigid sphere and the high-order reflections from the walls. The center of the coordinate system coincides with a bottom corner of the room. The dimensions of the room were 3.5 m wide (along x), 3 m long (along y), and 3 m high (along z). The reflection coefficients of all walls were 0.3. A maximum reflection order was obtained by setting the corresponding parameter to -1. The center position of the microphone array was $(2, 1.5, 1)$ m. The frontal direction was along the positive x-axis. These dimensions were used to make the time-domain results visible within a frame of 10 ms.

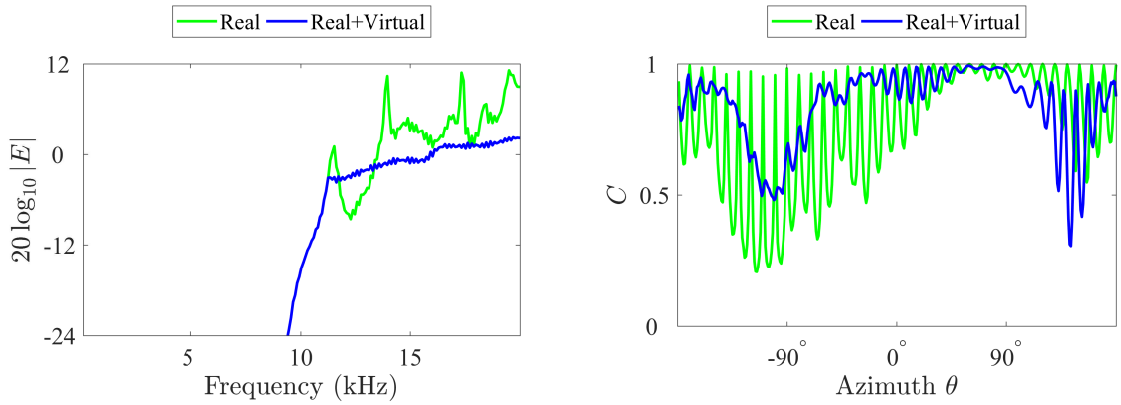
Figure 3 presents frequency-domain left-ear signals reconstructed in anechoic (top panels) and reverberant (bottom panels) conditions. Reconstructions for the *Target* condition $M_\mathrm{r} = 180$ are shown in panels (a) and (d). Reconstructions for the *Real* condition $M_\mathrm{r} = 36$ are shown in panels (b) and (e). It can be observed that using only $M_\mathrm{r} = 36$ real microphones is not enough to cover the higher frequencies. Reconstruction results for the *Real+Virtual* condition, when adding $M_\mathrm{v} = 144$ virtual microphones to the case of $M_\mathrm{r} = 36$ in the frequencies above $f_\mathrm{max} \approx 11$ Khz, are shown in panels (c) and (f). When comparing the results in (c) and (f) with the results in (b) and (e), respectively, it can be observed that adding virtual microphones enhances the performance at higher frequencies in both anechoic and reverberant conditions. Enhancement is mainly obtained on the side ipsilateral to the evaluated ear, as can be observed when comparing the high-frequency results for the azimuthal angles around $\theta = 90^\circ$.

Figure 4 shows time-domain left-ear signals reconstructed in anechoic (top panels) and reverberant (bottom panels) conditions. These signals correspond to the impulse responses of the transfer functions presented in Fig. 3. Similar to Fig. 3, reconstructions for the *Target*



(a) Error defined in (7). (b) Correlation coefficient defined in (8).

Fig. 5: Overall accuracy in anechoic conditions.



(a) Error defined in (7). (b) Correlation coefficient defined in (8).

Fig. 6: Overall accuracy in reverberant conditions.

condition are presented in panels (a) and (d); *Real* condition in (b) and (e); and *Virtual* condition in (c) and (f). A comparison of these results reveals that using only $M_\mathrm{r} = 36$ real microphones produces impulse responses with artifacts in the form of forward and backward replicas of the main peaks. These time-domain artifacts are attenuated when adding $M_\mathrm{v} = 144$ virtual microphones to the case of $M_\mathrm{r} = 36$. Adding virtual microphones in the high-frequency range thus enhances the temporal localization of the main peaks of the impulse responses in both anechoic and reverberant conditions.

Finally, overall reconstruction accuracies with respect to the *Target* condition, assuming anechoic and reverberant environments, were obtained by comparing the binaural signals $\hat{\boldsymbol{b}}$ for the *Real* or *Real+Virtual* conditions with the binaural signals $\boldsymbol{b}$ for the *Target* condition. Overall accuracies were evaluated in terms of magnitude deviations of transfer functions and similarities of impulse responses.

Magnitude deviations were calculated based on the following error:

$$E(f) = \frac{\mathrm{RMS}_\theta\left[\boldsymbol{b}(\theta,f) - \hat{\boldsymbol{b}}(\theta,f)\right]}{\mathrm{RMS}_\theta\left[\boldsymbol{b}(\theta,f)\right]}, \qquad (7)$$

where RMS_θ denotes the root mean square value along the azimuthal angle θ.

Similarities were calculated based on the correlation coefficient:

$$C(\theta) = \frac{\mathrm{Cov}_t\left[\mathrm{IR}\{\boldsymbol{b}\}(\theta,t), \mathrm{IR}\{\hat{\boldsymbol{b}}\}(\theta,t)\right]}{\left(\mathrm{Var}_t\left[\mathrm{IR}\{\boldsymbol{b}\}(\theta,t)\right]\mathrm{Var}_t\left[\mathrm{IR}\{\hat{\boldsymbol{b}}\}(\theta,t)\right]\right)^{\frac{1}{2}}}, \tag{8}$$

where IR denotes impulse responses, and Var_t and Cov_t denote variance and covariance along t, respectively.

Figs. 5 and 6 present the overall accuracies in anechoic and reverberant conditions, respectively. In both conditions, the results reveal that adding virtual microphones in the high-frequency range enhances the responses thanks to decreasing the overall magnitude deviations. In the time-domain, the benefits of adding virtual microphones are translated into higher similarities for sound source positions in front of the listener, that is, around the reference position $\theta = 0°$.

5 Limitations and possibilities

The binaural reconstruction framework used in the assessment of virtual microphone generation was restricted to the beamforming approach reported in [10, 11, 15]. However, since virtual microphone generation aims at increasing the number of recordings for its subsequent combination with HRTFs, the method can also be applied in other frameworks such as binaural ambisonics [9]. In general, virtual microphone generation constitutes a preprocessing stage intended to be applied before any array signal encoding is performed.

Recording environments with static conditions were assumed throughout evaluations. Conference rooms or concert halls can be roughly described by these conditions provided that listeners tend to remain still in their seats and sound sources are often confined to a small region in front of them. Nevertheless, when aiming at binaural reconstruction in real life situations, the reference position required for virtual microphone generation must be estimated during an actual recording. Addressing this problem would require an additional processing stage for sound source localization in real time. The use of methods such as the one proposed in [29] would be beneficial in this regard.

The positions of sources and transducers were restricted to the horizontal plane throughout evaluations. Given the rotational symmetry of the rigid sphere model, the results presented in this paper for a circular microphone array might be regarded as a representative case of spherical microphone arrays. Yet the assessment of a complete three-dimensional binaural reconstruction would necessarily require to consider spherical microphone arrays and spherical HRTF datasets in future evaluations.

6 Conclusion

The virtual microphone generation method introduced in [25] was assessed numerically in the context of binaural reconstruction from rigid circular microphone array recordings performed in anechoic and reverberant conditions. The results demonstrated that adding virtual microphones extends the frequency range of operation and attenuates time-domain artifacts.

Extensions of this work may include further energy-based criteria when selecting the nearest real microphone, as well as the combination of various numbers of nearby microphones to generate a virtual microphone. Additionally, perceptual evaluations by means of detectability of differences and localization tests along angles and distances are necessary to provide further insight into the validity of the suggested approach.

7 Acknowledgment

This work was supported by the JSPS Grant-in-Aid for Scientific Research (KAKENHI) under Grants JP16H01736 and JP17K12708.

References

[1] Altinsoy, M. E., "The quality of auditory-tactile virtual environments," *J. Audio Eng. Soc*, 60(1/2), pp. 38–46, 2012.

[2] Sakamoto, S., Hasegawa, G., Honda, A., Iwaya, Y., Suzuki, Y., and Gyoba, J., "Body vibration effects on perceived reality with multi-modal contents," *ITE Transactions on Media Technology and Applications*, 2(1), pp. 46–50, 2014, doi: 10.3169/mta.2.46.

[3] Steed, A. and Schroeder, R., "Collaboration in immersive and non-immersive virtual environments," in M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and J. R. Schaevitz, editors, *Immersed in Media: Telepresence Theory, Measurement &*

Technology, pp. 263–282, Springer International Publishing, Cham, 2015, ISBN 978-3-319-10190-3.

[4] Hendrix, C. and Barfield, W., "The sense of presence within auditory virtual environments," *Presence: Teleop. Virt. Env.*, 5(3), pp. 290–301, 1996, ISSN 1054-7460, doi:10.1162/pres.1996.5.3.290.

[5] Algazi, V. and Duda, R., "Headphone-Based Spatial Sound," *IEEE Signal Process. Mag.*, 28(1), pp. 33–42, 2011, ISSN 1053-5888, doi:10.1109/MSP.2010.938756.

[6] Betlehem, T., Zhang, W., Poletti, M., and Abhayapala, T., "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, 32(2), pp. 81–91, 2015, ISSN 1053-5888, doi:10.1109/MSP.2014.2360707.

[7] 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, doi:10.1121/1.4795780.

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

[9] 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.

[10] 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, ISSN 1745-1337.

[11] 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, doi:10.1250/ast.36.516.

[12] 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, doi:10.1121/1.4800833.

[13] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Numerical evaluation of binaural synthesis from rigid spherical microphone array recordings," in *Proc. Audio Eng. Soc. Int. Conf. Headphone Technology*, Audio Engineering Society, Aalborg, Denmark, 2016, ISBN 978-1-942220-09-1, doi:10.17743/aesconf.2016.978-1-942220-09-1.

[14] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Spatial accuracy of binaural synthesis from rigid spherical microphone array recordings," *Acoust. Sci. Technol.*, 38(1), pp. 23–30, 2017, doi:10.1250/ast.38.23.

[15] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets," *Acoust. Sci. Technol.*, 38(2), pp. 51–62, 2017, doi:10.1250/ast.38.51.

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

[17] Pollow, M., Nguyen, K.-V., Warusfel, O., Carpentier, T., Müller-Trapet, M., Vorländer, M., and Noisternig, M., "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics," *Acta Acust. United Ac.*, 98(1), pp. 72–82, 2012, doi:10.3813/AAA.918493.

[18] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values," *Acoust. Sci. Technol.*, 38(1), pp. 1–13, 2017, doi:10.1250/ast.38.1.

[19] 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, doi:10.1121/1.2191608.

[20] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Dataset of near-distance head-related

transfer functions calculated using the boundary element method," Audio Engineering Society, Tokyo, Japan, 2018, to be presented.

[21] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Boundary matching filters for spherical microphone and loudspeaker arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 26(3), pp. 461–474, 2018, ISSN 2329-9304, doi:10.1109/TASLP.2017.2778562.

[22] Algazi, V. R., Duda, R. O., and Hom, R. C.-M., "High-frequency interpolation for motion-tracked binaural sound," in *Proc. 121th Convention Audio Eng. Soc.*, 2006.

[23] Meier, M., Weitnauer, M., and Groh, J., "Spherical surface microphone interpolation," in *Proc. Int. Conf. Spatial Audio*, 2011.

[24] 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.

[25] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., "Enhancement of spatial sound recordings by adding virtual microphones to spherical microphone arrays," *J. Inf. Hiding and Multimedia. Signal Process.*, 8(6), pp. 1392–1404, 2017, ISSN 20734212.

[26] Ajdler, T., Sbaiz, L., and Vetterli, M., "Plenacoustic function on the circle with application to HRTF interpolation," in *Proc. IEEE ICASSP*, volume 3, pp. 273–276, 2005, doi:10.1109/ICASSP.2005.1415699.

[27] Ajdler, T., Faller, C., Sbaiz, L., and Vetterli, M., "Sound field analysis along a circle and its applications to HRTF interpolation," *J. Audio Eng. Soc.*, 56(3), pp. 156–175, 2008.

[28] Jarret, D. P., Habets, E. A. P., Thomas, M. R. P., and Naylor, P. A., "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acoust. Soc. Am.*, 132(3), pp. 1462–1472, 2012, doi:10.1121/1.4740497.

[29] Pavlidi, D., Griffin, A., Puigt, M., and Mouchtaris, A., "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Trans. Audio, Speech, Language Process.*, 21(10), pp. 2193–2206, 2013, ISSN 1558-7916, doi:10.1109/TASL.2013.2272524.