Audio Engineering Society

Conference Paper

Presented at the 2020 Conference on
Audio for Virtual and Augmented Reality
2020 August 17 – 19, Redmond, WA

Beamforming-based Binaural Reproduction by Matching of Binaural Signals

Lior Madmoni[1], Jacob Donley[2], Vladimir Tourbabin[2], and Boaz Rafaely[1]

[1]*School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel*
[2]*Facebook Reality Labs, Redmond, WA, USA*

Correspondence should be addressed to Lior Madmoni (`liormad@post.bgu.ac.il`)

ABSTRACT

The capture and reproduction of spatial audio is becoming increasingly popular, with the mushrooming of applications in teleconferencing, entertainment and virtual reality. Two popular methods include high order Ambisonics and beamforming-based binaural reproduction, which are mainly studied with spherical arrays. However, a design framework for arrays with a less regular configuration has yet to be formulated and studied. This paper studies a method for binaural reproduction with general arrays, based on the direct estimation of the binaural signal from the array measurements. A simulation study shows that this method is capable of performing binaural reproduction with high accuracy over a limited frequency range, even with a small six-microphone array.

1 Introduction

Binaural reproduction has recently become a popular topic of research [1, 2, 3], with the rapid growth of many applications including entertainment, auditory research and virtual reality. A common method for the rendering of binaural signals uses high order Ambisonics (HOA) signals [4, 5], together with head-related transfer functions (HRTFs) [6]. The HOA signals are typically synthesized with a computer or estimated from microphone-array measurements [7, 8, 9]. An accurate estimation of HOA signals is often required in order to render binaural signals with good perceptual qualities. Hence, many research works studied methods for capturing HOA signals [10], and their incorporation into binaural reproduction [11, 6]. Such methods are most suitable for arrays with high directional resolution over the entire 3-D space, and hence, spherical arrays are typically used.

Another approach which is not limited to spherical arrays and which is increasing in popularity is beamforming-based binaural reproduction (BFBR) [12, 13]. With this approach, a set of beamformers is applied to the microphone signals with different look-directions, followed by convolution with the HRTFs from the same directions. Finally, the output signals are weighted and summed to produce the binaural signals. BFBR is more general than HOA, but to date most published studies have used spherical arrays

[12, 13, 14, 15]. Moreover, these works did not provide a comprehensive theoretical framework for BFBR nor guidelines for selecting the number of beamformers, look-directions or beamformer types. However, a preliminary work on one such framework was recently developed in [16]. This work presents well-defined conditions for the equivalence between BFBR and HOA with spherical arrays, which will be briefly summarized in this paper. While these conditions facilitate the design of BFBR systems, it should be noted that they are limited to spherical arrays.

While spherical arrays are indeed capable of providing high quality binaural reproduction, a more flexible array geometry may be desired. Some examples include microphones on handheld devices or wearable arrays, which are typically limited by the geometry of the device on which they are mounted. In addition, compact arrays with a small number of microphones may also be desired due to limited space or computational power. For such arrays, the beam-pattern may depend on the look-direction, and their directional resolution may be limited, such that they are not capable of capturing HOA signals. Hence, performing a high quality binaural reproduction with measurements that have been captured with general arrays is still a great challenge.

This paper presents a method for binaural reproduction which is suitable for a general array geometry, aiming to overcome current limitations of BFBR design for general arrays. The method is based on estimating the binaural signal directly from the array measurements with minimal error, under the assumption of a sound field which is comprised of uncorrelated sources. While similar methods that estimate the binaural signals were presented previously in [17, 18, 19, 20], the accuracy of the estimated binaural signals was not studied in detail for general arrays.

The paper is organized as follows: Section 2 provides a mathematical background to array processing and binaural reproduction. Section 3 summarizes the previous work for BFBR design, that is described in [16]. In Section 4, the proposed method for binaural signals estimation with general arrays is developed, and the accuracy of this method is analyzed in Section 5 with a simulation study. The results from this analysis are discussed in Section 6, which also concludes the paper.

2 Background

This section provides mathematical background concerning array processing and binaural reproduction, using models that are used in the following sections. Throughout the paper, the spherical coordinate system will be used, denoted by (r,θ,ϕ), where r is the distance from the origin, θ is the elevation angle measured from the positive z axis downward to the xy plane, and ϕ is the azimuth angle measured from the positive x axis to the positive y axis.

2.1 Array measurement model

Assume a sound field that can be described by a directional plane-wave density (PWD) function $a(k,\theta,\phi)$, where $k=\frac{2\pi}{\lambda}$ is the wave-number and λ is the wavelength. Further assume that the sound field is comprised of L far-field sound sources arriving from directions $\{(\theta_l,\phi_l)\}_{l=1}^{L}$ corresponding to the signals $\{s_l(k)\}_{l=1}^{L}$. The sound field is captured by an array with M microphones which are located at $\{(r_m,\theta_m,\phi_m)\}_{m=1}^{M}$, centered at the origin. The noisy array measurements can be described by the following narrow-band model [21]:

$$\mathbf{x}(k)=\mathbf{V}(k)\mathbf{s}(k)+\mathbf{n}(k), \tag{1}$$

where $\mathbf{x}(k)=\begin{bmatrix}x_1(k), & x_2(k), & \ldots, & x_M(k)\end{bmatrix}$ is the microphone signals (measurements) vector, $\mathbf{V}(k)=\begin{bmatrix}\mathbf{v}(k,\theta_1,\phi_1), & \mathbf{v}(k,\theta_2,\phi_2), & \ldots, & \mathbf{v}(k,\theta_L,\phi_L)\end{bmatrix}$ is an $M\times L$ complex matrix with columns $\mathbf{v}(k,\theta_l,\phi_l)$ representing the array steering vector from the l-th source to the microphone positions for all $l=1,2,...,L$ [21], $\mathbf{s}(k)=\begin{bmatrix}s_1(k), & s_2(k), & \ldots, & s_L(k)\end{bmatrix}$ is the source signals vector, and $\mathbf{n}(k)$ is an additive noise vector.

2.2 Beamforming

Beamforming is a method for spatial filtering of array measurements that is applied in order to produce a desired output. Examples for such outputs include undistorted, enhanced or attenuated responses at desired directions. It is performed by multiplying the array measurements with the beamformer weights, followed by summation [21]:

$$y(k,\theta_d,\phi_d)=\mathbf{w}_d^H(k)\mathbf{x}(k), \tag{2}$$

where $y(k,\theta_d,\phi_d)$ is the beamformer output for look-direction (θ_d,ϕ_d), $\mathbf{w}_d(k)$ is an $M\times 1$ complex vector of the beamformer weights and $(\cdot)^H$ is the Hermitian operator.

An important measure of a beamformer is the directivity factor (DF). It is defined as the power of the beamformer output due to a single unity-amplitude plane-wave input from the look-direction (θ_d, ϕ_d), normalized by the average power over all directions [21]:

$$DF(k, \theta_d, \phi_d) = 4\pi \frac{|\mathbf{w}_d^H(k)\mathbf{v}(k,\theta_d,\phi_d)|^2}{\int_0^{2\pi}\int_0^{\pi} |\mathbf{w}_d^H(k)\mathbf{v}(k,\theta,\phi)|^2 \sin\theta d\theta d\phi}, \tag{3}$$

where $\mathbf{v}(k, \theta, \phi)$ are the steering vectors of the array, defined similarly to in (1). This measure quantifies the spatial resolution of the beamformer, since a very large $DF(k, \theta_d, \phi_d)$ means that the beamformer can capture isolated spatial information from (θ_d, ϕ_d), with very low contributions from other nearby directions.

2.3 Binaural Reproduction

A sound field which is described by the PWD function $a(k, \theta, \phi)$ can be used for binaural reproduction, by integration with the HRTFs of the left and right ears [6]:

$$p^{l,r}(k) = \int_0^{2\pi}\int_0^{\pi} a(k, \theta, \phi) h^{l,r}(k, \theta, \phi) \sin\theta d\theta d\phi, \tag{4}$$

where $h^{l,r}(k, \theta, \phi)$ are the HRTFs of the left and right ears, denoted by $(\cdot)^l$ and $(\cdot)^r$, respectively. Alternatively, (4) can be formulated in the spherical harmonics (SH) domain [6]:

$$p^{l,r}(k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \tilde{a}_{nm}^*(k) h_{nm}^{l,r}(k), \tag{5}$$

where n and m are the SH order and degree [7], respectively, $\tilde{a}_{nm}(k)$ are the spherical Fourier transform (SFT) [7] of the complex conjugate of the PWD coefficients, $h_{nm}^{l,r}(k)$ are the SFT coefficients of the left and right HRTFs, and $(\cdot)^*$ denotes the complex conjugate.

In practice, the sound field is measured by an array, such that the PWD function can be calculated from the array measurements with low error up to a finite SH order, denoted by N_a [7]. In addition, the HRTFs are also typically numerically calculated or measured up to a finite SH order, denoted by N_h [22]. Hence, the infinite summation in (5) can be truncated with low error to a finite order $N_p = \min\{N_a, N_h\}$:

$$p^{l,r}(k) = \sum_{n=0}^{N_p}\sum_{m=-n}^{n} \tilde{a}_{nm}^*(k) h_{nm}^{l,r}(k). \tag{6}$$

3 Previous Work: Beamforming-Based Binaural Reproduction

This section presents the beamforming-based approach for binaural reproduction. First, the method is presented generally, followed by a summary of a recent work that presents a theoretical framework for the design of BFBR systems [16]. In BFBR, the outputs from several beamformers are weighted and linearly combined with the HRTFs from the corresponding look-directions:

$$\hat{p}^{l,r}(k) = \sum_{d=1}^{D} \alpha_d h^{l,r}(k, \theta_d, \phi_d) y(k, \theta_d, \phi_d), \tag{7}$$

where $\{\alpha_d\}_{d=1}^D$ are the design weights and $y(k, \theta_d, \phi_d)$ is the beamformer output as in (2). The primary objective of BFBR design is to appropriately select the weights, the beam-patterns, and the set of directions $\{(\theta_d, \phi_d)\}_{d=1}^D$ in (7).

When a spherical array is used to produce $y(k, \theta_d, \phi_d)$, as in (7), with a maximum directivity beamformer [21], the array output is equivalent to plane-wave decomposition from the look-direction [7, 16, 23]. Under these assumptions, and based on the work in [24], it was shown in [16] that choosing the look-directions $\{(\theta_d, \phi_d)\}_{d=1}^D$ according to a sampling scheme on the sphere that is designed to be aliasing-free for SH orders of $\max\{N_a, N_h\}$ and below, and choosing $\{\alpha_d\}_{d=1}^D$ according to the corresponding sampling weights, will yield an equivalence between (7) and the HOA representation in (6), which is order limited to $N_p = \min\{N_a, N_h\}$. Hence, the use of maximum directivity beamformers together with this suggested sampling scheme in (7) defines the conditions for the equivalence between BFBR and HOA based reproduction [16]. In addition, it was shown in [16] that, for the conditions above to hold, the number of steering directions, D, should also be greater than or equal to the DF of the maximum directivity beamformer. This provides a simple guideline for selecting D, since the beam-pattern of the maximum directivity beamformer with spherical arrays is independent of the look-direction and frequency. The conditions and relations presented above provide a framework for the design of a BFBR system with spherical arrays.

Contrary to spherical arrays, the DF of the maximum directivity beamformer of general arrays may depend on the look-direction and on the frequency. Hence, the guideline for choosing D, based on the DF of the maximum directivity beamformer, is not suitable for non-spherical arrays. Nonetheless, [16] suggests a way to extend this guideline to general arrays, by using the average DF over all look-directions. Note that with general arrays, D may also be frequency dependent.

This extension for general arrays assumes that a maximum directivity beamformer is used with any general array for BFBR. However, it does not provide a method or guidelines for selecting the correpsonding steering directions $\{(\theta_d,\phi_d)\}_{d=1}^{D}$ or the weights $\{\alpha_d\}_{d=1}^{D}$ in (7). The following section aims to overcome some of these limitations, by providing an alternative approach for binaural reproduction with general arrays.

4 The Proposed Method For Binaural Reproduction with General Arrays

This section presents a different approach to array-based binaural reproduction - by estimating the binaural signals in (4) with minimal error. For simplicity, the dependency on the wave-number k is omitted henceforth, while all of the following derivations are for narrow-band system models [21].

Following from the assumptions in subsection 2.1, since the sound field is comprised of L sound sources, (4) can be reduced to

$$p^{l,r} = [\mathbf{h}^{l,r}]^T \mathbf{s}, \tag{8}$$

where

$$\mathbf{h}^{l,r} = \left[h^{l,r}(\theta_1,\phi_1), \quad h^{l,r}(\theta_2,\phi_2), \quad \ldots, \quad h^{l,r}(\theta_L,\phi_L) \right]$$

contains the HRTFs corresponding to the directions of the sources. Next, assume that a general array with a known configuration is used to capture this sound field, with array measurements according to the model in (1), such that the steering matrix $\mathbf{V}(k)$ can be calculated analytically or numerically, or it can be measured. In the first step, the array measurements are filtered and combined, in a similar manner to beamforming in (2):

$$z = \mathbf{c}^H \mathbf{x}, \tag{9}$$

where $\mathbf{c}$ is an $M\times 1$ vector holding the filter coefficients. Next, $\mathbf{c}$ is chosen to minimize the following mean-squared error between (9) and $p^{l,r}$, the binaural signals in (8), for each ear separately:

$$err_{\text{bin}}^{l,r} = \mathbb{E}\{|p^{l,r} - z^{l,r}|^2\}, \tag{10}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator and $z^{l,r}$ is defined as in (9) with $\mathbf{c}^{l,r}$, which correspond to the binaural signals of the left and right ears, respectively. Next, assume that the sources are spatially white with a covariance matrix $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \sigma_s^2 \mathbf{I}_L$, where $\mathbf{I}_L$ is the identity matrix of size L, and that the noise is uncorrelated to the sources and is also white with a covariance matrix $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_M$. Then, substituting (8) and (9) in (10) leads to the following error formulation:

$$err_{\text{bin}}^{l,r} = \sigma_s^2 \left\| \mathbf{V}^H \mathbf{c}^{l,r} - [\mathbf{h}^{l,r}]^* \right\|_2^2 + \sigma_n^2 \left\| \mathbf{c}^{l,r} \right\|_2^2, \tag{11}$$

where $\|\cdot\|_2$ is the l^2-norm. In order to produce an accurate binaural signal, (11) is minimized over the filter $\mathbf{c}^{l,r}$ for each ear, and it can be shown that the optimal filter can be formulated as [25]:

$$\mathbf{c}_{\text{TR}}^{l,r} = (\mathbf{V}\mathbf{V}^H + \frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}_M)^{-1}\mathbf{V}[\mathbf{h}^{l,r}]^*, \tag{12}$$

where $\mathbf{c}_{\text{TR}}^{l,r}$ are the optimal filters for the left and right ears, respectively; this coincides with the Tikhonov-regularization method. Finally, binaural reproduction can be performed by substituting (12) in (9):

$$\hat{p}_{\text{mic-TR}}^{l,r} = [\mathbf{c}_{\text{TR}}^{l,r}]^H \mathbf{x} = [\mathbf{h}^{l,r}]^T \mathbf{V}^H (\mathbf{V}\mathbf{V}^H + \frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}_M)^{-1}\mathbf{x}, \tag{13}$$

where $\hat{p}_{\text{mic-TR}}^{l,r}$ is the estimated binaural signal from the microphone signals. Similar methods that minimize the error in (11) were proposed in [17, 18, 19, 20]. However, these mainly studied spherical arrays, rather then general arrays, and the quality of the reproduced sound has not been studied in detail.

In order to study the performance of the proposed method in (13), the following normalized error is defined, using (8) and (11):

$$\begin{aligned} err_{\text{bin-norm}}^{l,r} &= \frac{\mathbb{E}\{|p^{l,r} - \hat{p}_{\text{mic-TR}}^{l,r}|^2\}}{\mathbb{E}\{|p^{l,r}|^2\}} \\ &= \frac{\sigma_s^2 \left\| \mathbf{V}^H \mathbf{c}_{\text{TR}}^{l,r} - [\mathbf{h}^{l,r}]^* \right\|_2^2 + \sigma_n^2 \left\| \mathbf{c}_{\text{TR}}^{l,r} \right\|_2^2}{\sigma_s^2 \left\| [\mathbf{h}^{l,r}]^* \right\|_2^2}. \end{aligned} \tag{14}$$

Note that if the signal-to-noise ratio (SNR) is high, such that $\sigma_s^2 \gg \sigma_n^2$, (14) reduces to

$$err_{\text{bin-norm}}^{l,r} \approx \frac{\left\| \mathbf{V}^H \mathbf{c}_{\text{TR}}^{l,r} - [\mathbf{h}^{l,r}]^* \right\|_2^2}{\left\| [\mathbf{h}^{l,r}]^* \right\|_2^2}, \tag{15}$$

which is equivalent to the normalized error of the HRTFs estimation from the array steering vectors. This may suggest that when the array configuration is somewhat similar to a head-like shape (for instance, a microphone array mounted on an approximately spherical body), the estimation error of the HRTFs from the steering vectors may be small. Hence, the estimation error of the binaural signal may also be small.

Recall that the proposed method assumes a sound-field which is comprised of L uncorrelated sound sources. At first, this may appear to be a limitation of the proposed method. However, if L is chosen to be sufficiently large, and the error in (14) is very small, it implies that the proposed method may be capable of high quality binaural reproduction for many other types of sound fields. This is because a sound field that is comprised of many uncorrelated sources is very difficult to reproduce, since each component is estimated separately and independently from the other components. Furthermore, this method is readily modifiable to other sound field models, by substituting the corresponding covariance matrix of the sound sources when calculating the expectation in (10) with (8) and (9).

5 Simulation Study

This section presents a simulation study of the proposed method for binaural reproduction with general arrays, that was described in Section 4. The study includes an error analysis of the estimated left binaural signal, as described in (14). Only the error corresponding to the left ear is studied, for simplicity. Recall that the proposed method was developed assuming a sound field which is comprised of L far-field sources. Hence, the error in (14) is studied for various values of L, specifically with 6, 20, 32 and 240 sources that arrive from directions with a spiral nearly-uniform distribution on the sphere [26]. This range for L was chosen, since with a large number of sources ($L = 240$), an accurate binaural signal estimation means that the proposed method is capable of high quality binaural reproduction for a relatively complex environment, even when the sources are uncorrelated. With lower values of L, the method is somewhat similar to BFBR, which produces binaural signals from a small set of look-directions.

Two different arrays are studied here:

1. A rigid spherical array with a radius of $r = 4.2\,\text{cm}$ and $M = 32$ microphones with a nearly-uniform configuration according to [27].
2. An array with $M = 6$ microphones that are uniformly distributed on a semi-circle in the horizontal plane, where $\{\theta_m = \frac{\pi}{2}\}_{m=1}^{M}$ and $\{\phi_m = \frac{\pi}{2} - \frac{\pi(m-1)}{M-1}\}_{m=1}^{M}$, and which are mounted on a rigid sphere with a radius of $r = 10\,\text{cm}$. This array contains many fewer microphones compared to the spherical array, and is an example of an array which can be mounted on a wearable device, and which is incapable of capturing HOA signals.

The steering vectors of both rigid arrays were calculated in the SH domain, as described in [7] (section 4.2), and according to the corresponding number of sources L. For the HRTFs in (14), the Neumann KU100 manikin measurements from the Cologne database [28] were used, with directions corresponding to the L sound sources. This was performed by first calculating the HRTFs in the SH domain up to an order of $N_h = 30$, and then applying the inverse spherical Fourier transform at the corresponding directions of the L sources. Next, the optimal filter $\mathbf{c}_{\text{TR}}^{l}$ was calculated according to (12) with an assumed SNR of 20 dB, for both arrays. Finally, the normalized error of the left ear in (14) was calculated for frequencies in the range [75, 24000] Hz, with steps of 75 Hz, also for both arrays.

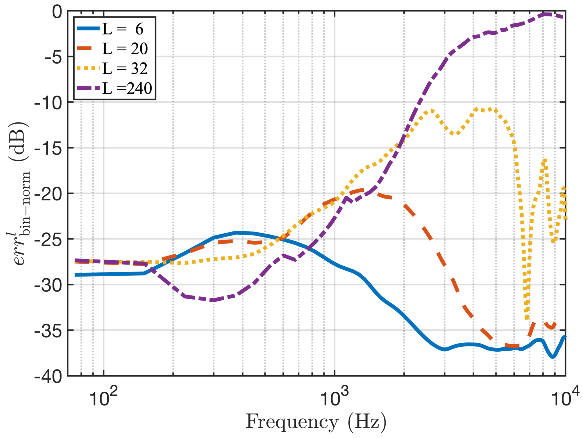
The normalized error that corresponds to the spherical array is presented in Fig. 1(a). Note the small error (below -15 dB) for the cases of 6 and 20 sound sources, over the entire frequency range. When there are 32 sources, the error is generally larger, but still lower than -10 dB. Finally, for the case of 240 sources, a small error is achieved at frequencies below 2 kHz, but at higher frequencies the error is significantly larger, such that the estimation of the binaural signal is no longer accurate. This increase in the error as L increases is expected, since it is generally more challenging to produce an accurate binaural signal when the sound field is comprised of more uncorrelated sources, as explained before.

Next, the normalized error of the estimated binaural signal with the semi-circular array is presented in Fig. 1(b). Note that when there are 6 sources, the error is very small over a relatively wide frequency range of 75-7000 Hz. This means that the array is capable of a relatively accurate reproduction of the binaural signal, when it is comprised of a small number of sources in this frequency range. In addition, the errors are very similar when comparing the cases of 20, 32 or 240 sound sources. More importantly, the binaural estimation is relatively accurate in these cases for frequencies lower than 1.5 kHz, as well. Above this frequency, the error increases significantly, such that an accurate binaural reproduction is no longer possible. Furthermore, note that below 1.5 kHz, the estimation is relatively accurate, even though the sources are not confined to the horizontal plane. This may be explained by the array steering vectors, which may have some similarity to the HRTFs, since they are measured around a sphere with a radius of 10 cm (which is somewhat similar to a head), as explained in Section 4 (see (15)). This means that the very simple semi-circular array may be capable of high quality binaural reproduction over a limited frequency range, even in complex acoustic environments, without any prior information.

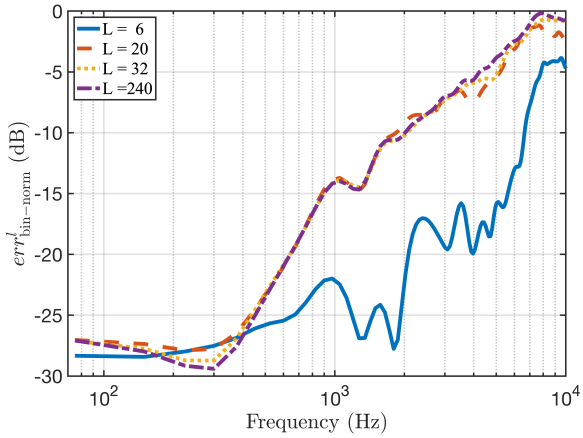
In addition to the objective analysis described above, an informal listening test was conducted. In this test, both speech and castanets sound signals were studied separately. Each signal was simulated as a point source in a room of dimensions $15.5 \times 9.8 \times 7.5$ m and with a reverberation time of $T_{60} = 1$ s. The recordings from the spherical and semi-circular arrays, positioned approximately 4 m away from each source, were then simulated. The same filters $\mathbf{c}^{l}_{\mathrm{TR}}$ as previously described were used to estimate the binaural signals, with $L = 6, 20, 32, 240$. These binaural signals were compared to a HOA signal, truncated to order $N_p = 30$, as described in (6). It was reported that as L increases, the spatial attributes of the estimated signal are more similar to the reference signal, in addition to the introduction of low-pass-filter effects, as was similarly reported in [16]. This was reported for both the speech and castanets signals and with both arrays. However, a formal and more elaborate subjective study is left for future work.

6 Discussion and Conclusions

This paper studied a method for binaural reproduction with general arrays. The method is based on estimating the binaural signal directly from the array measurements, under the assumptions of a sound field which is comprised of a discrete set of uncorrelated sound sources. A simulation study showed that with both a spherical array and a semi-circular array, an accurate estimation of the binaural signal is possible over a limited frequency range, and for a large number of sources with directions that are distributed nearly-uniformly in 3-D space. While the estimation with the spherical



(a) Spherical array



(b) Semi-circular array

Fig. 1: The normalized error of the estimated left binaural signal described in (14) as a function of frequency, and for various numbers of sources L. (a) estimated with a spherical array, (b) estimated with a semi-circular array.

array is much more accurate, an array with many fewer microphones may be desired for some applications, where space or computational power are scarce. Most importantly, the semi-circular array has the potential to perform high quality binaural reproduction with complex sound fields that are comprised of sound sources arriving from directions which are not limited to the horizontal plane. For future work, it is suggested to further study this method with different databases of HRTFs and other sound field models, and to perform an elaborate subjective study with listening tests to further verify these results.

Acknowledgment

This research was supported by Facebook Reality Labs.

References

[1] Politis, A., McCormack, L., and Pulkki, V., "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 379–383, IEEE, 2017.

[2] Noisternig, M., Sontacchi, A., Musil, T., and Holdrich, R., "A 3D ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, Audio Engineering Society, 2003.

[3] Ben-Hur, Z., Brinkmann, F., Sheaffer, J., Weinzierl, S., and Rafaely, B., "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *The Journal of the Acoustical Society of America*, 141(6), pp. 4087–4096, 2017.

[4] Rafaely, B. and Avni, A., "Interaural cross correlation in a sound field represented by spherical harmonics," *The Journal of the Acoustical Society of America*, 127(2), pp. 823–828, 2010.

[5] Poletti, M. A., "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, 53(11), pp. 1004–1025, 2005.

[6] 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," *The Journal of the Acoustical Society of America*, 133(5), pp. 2711–2721, 2013.

[7] Rafaely, B., *Fundamentals of spherical array processing*, volume 8, Springer, 2015.

[8] Wabnitz, A., Epain, N., McEwan, A., and Jin, C., "Upscaling ambisonic sound scenes using compressed sensing techniques," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 1–4, IEEE, 2011.

[9] Alon, D. L., Sheaffer, J., and Rafaely, B., "Robust plane-wave decomposition of spherical microphone array recordings for binaural sound reproduction," *The Journal of the Acoustical Society of America*, 138(3), pp. 1925–1926, 2015.

[10] Abhayapala, T. D. and Ward, D. B., "Theory and design of high order sound field microphones using spherical microphone array," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pp. II–1949, IEEE, 2002.

[11] Bernschütz, B., Giner, A. V., Pörschmann, C., and Arend, J., "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, 100(5), pp. 972–983, 2014.

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

[13] O'Donovan, A. M., Zotkin, D. N., and Duraiswami, R., "Spherical microphone array based immersive audio scene rendering," International Community for Auditory Display, 2008.

[14] Song, W., Ellermeier, W., and Hald, J., "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *The Journal of the Acoustical Society of America*, 123(2), pp. 910–924, 2008.

[15] Jiang, J., Xie, B., and Mai, H., "The Number of Virtual Loudspeakers and the Error for Spherical Microphone Array Recording and Binaural Rendering," in *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*, Audio Engineering Society, 2018.

[16] Ifergan, I. and Rafaely, B., "Theoretical Framework for Beamformer Distribution in Beamforming based Binaural Reproduction," *MS.c. Thesis, School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, In submission*, 2020.

[17] Salvador, C. D., Sakamoto, S., Trevino, J., and Suzuki, Y., "Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets," *Acoustical Science and Technology*, 38(2), pp. 51–62, 2017.

[18] Poletti, M. A. and Svensson, U. P., "Beamforming synthesis of binaural responses from computer simulations of acoustic spaces," *The Journal of the Acoustical Society of America*, 124(1), pp. 301–315, 2008.

[19] Calamia, P., Davis, S., Smalt, C., and Weston, C., "A conformal, helmet-mounted microphone array for auditory situational awareness and hearing protection," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 96–100, 2017.

[20] Zhao, S., Rogowski, R., Johnson, R., and Jones, D. L., "3D binaural audio capture and reproduction using a miniature microphone array," in *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx)*, pp. 151–154, 2012.

[21] Van Trees, H. L., *Optimum array processing: Part IV of detection, estimation, and modulation theory*, John Wiley & Sons, 2004.

[22] Evans, M. J., Angus, J. A., and Tew, A. I., "Analyzing head-related transfer function measurements using surface spherical harmonics," *The Journal of the Acoustical Society of America*, 104(4), pp. 2400–2411, 1998.

[23] Meyer, J. and Elko, G., "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pp. II–1781–II–1784, 2002.

[24] Ben-Hur, Z., Sheaffer, J., and Rafaely, B., "Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction," *Applied Acoustics*, 134, pp. 138–144, 2018.

[25] Tikhonov, A. N., Goncharsky, A., Stepanov, V., and Yagola, A. G., *Numerical methods for the solution of ill-posed problems*, volume 328, Springer Science & Business Media, 2013.

[26] Saff, E. B. and Kuijlaars, A. B., "Distributing many points on a sphere," *The mathematical intelligencer*, 19(1), pp. 5–11, 1997.

[27] Hardin, R. H. and Sloane, N. J., "McLaren's improved snub cube and other new spherical designs in three dimensions," *Discrete & Computational Geometry*, 15(4), pp. 429–441, 1996.

[28] Bernschütz, B., "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA) conference on acoustics*, p. 29, AIA/DAGA, 2013.