

RESEARCH

Open Access

# On the selection of the number of beamformers in beamforming-based binaural reproduction



Itay Ifergan\*  and Boaz Rafaely

## Abstract

In recent years, spatial audio reproduction has been widely researched with many studies focusing on headphone-based spatial reproduction. A popular format for spatial audio is higher order Ambisonics (HOA), where a spherical microphone array is typically used to obtain the HOA signals. When a spherical array is not available, beamforming-based binaural reproduction (BFBR) can be used, where signals are captured with arrays of a general configuration. While shown to be useful, no comprehensive studies of BFBR have been presented and so its limitations and other design aspects are not well understood. This paper takes an initial step towards developing a theory for BFBR and develops guidelines for selecting the number of beamformers. In particular, the *average directivity factor* of the microphone array is proposed as a measure for supporting this selection. The effect of head-related transfer function (HRTF) order truncation that occurs when using too many beamformer directions is presented and studied. In addition, the relation between HOA-based binaural reproduction and BFBR is discussed through analysis based on a spherical array. A simulation study is then presented, based on both a spherical and a planar array, demonstrating the proposed guidelines. A listening test verifies the perceptual attributes of the methods presented in this study. These results can be used for more informed beamformer design for BFBR.

**Keywords:** Binaural reproduction, Microphone arrays, Beamforming, Higher order Ambisonics, Directivity factor

## 1 Introduction

Spatial audio has become a common feature in many application areas, such as virtual reality, hearing aids, teleconferencing, and entertainment. These applications use spatial audio to improve speech intelligibility or to spatially reproduce an acoustic scene for an immersive auditory experience.

A common format for spatial sound reproduction is higher order Ambisonics (HOA), which has lately been incorporated into the MPEG coding standard [1]. This approach aims to reproduce the sound pressure at the ears by employing spherical harmonics (SH) representations and using headphones [2] or loudspeaker arrays

[3] for playback. The method became popular due to its convenient mathematical framework and its ability to capture HOA signals from a wide range of spherical microphone array configurations [4–7]. In addition, spatial audio recording and reproduction based on HOA has been well studied and its behavior and limitations are reasonably well understood [8–13]. HOA signals are also employed for binaural reproduction, where spherical microphone arrays [14, 15] are used to capture information about the sound field, which is then combined with head-related transfer function (HRTF) [16] representations in the SH domain. However, for spherical microphone arrays with a small number of microphones and a low SH order, the spatial resolution becomes poor and the quality of the spatial audio is degraded [13]. While many studies have been done on this subject [17–21], with the objective of improving the spectral coloration

\*Correspondence: [ifergani@post.bgu.ac.il](mailto:ifergani@post.bgu.ac.il)

School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva, Israel

due to SH order truncation, leading to improved perception, the spatial properties of the signal have not been significantly improved. Furthermore, the method is not applicable to those arrays of general configurations which cannot provide HOA signals.

Alternative approaches which do not require spherical arrays have recently been developed. One such approach is the virtual artificial head (VAH) [22, 23], designed for anechoic environments but was recently studied for reverberant environment [24], which is based on applying beamforming to the microphone signals leading to binaural reproduction. Another similar approach, called beamforming-based binaural reproduction (BFBR), has been studied in [25, 26]. This approach was originally applied for spherical microphone arrays by using beamforming for plane-wave decomposition (PWD) at selected directions, followed by convolving the beamformer outputs with HRTFs in the same directions. This is similar to the virtual loudspeaker approach which is typically employed on the HOA signal [15], whereas for BFBR the application of beamforming does not necessarily incorporate, or leads to HOA signals, although this may depend on the beamforming design and the microphone array used. Other approaches, similar to BFBR, were also previously investigated in [27, 28] in which loudspeaker encoding techniques were applied on the Ambisonics signal, typically captured from a B-Format microphone array [29], instead of beamforming. Several studies employing BFBR then followed. The authors of [30] performed a psychoacoustic evaluation of this approach in a study of target detection in noise, as an alternative to binaural recordings from dummy heads. In [31], their research was extended to investigate the use of BFBR in the psychoacoustic evaluation of multichannel loudspeaker array configurations. The authors of [32] applied BFBR with maximum white noise gain beamformers, using a helmet-mounted microphone array for hearing protection while maintaining auditory situational awareness. Recently, the authors of [33] studied the error of binaural reproduction when increasing the number of beamformers, using BFBR with spherical arrays. Their results suggested that increasing the number of beamformers leads to a decrease in reproduction error and that this effect is saturated at some number of beamformers. However, their results are observational only and limited to spherical arrays. Furthermore, no psychoacoustical evaluation has been presented to support this claim. Another relevant recent study on binaural reproduction with general array configurations was presented in [34], where the authors proposed a design based on the matching of the array output to binaural signals. However, the work only discussed synthetic sound fields with a finite number of incoherent sources and was not applied or studied with realistic sound fields, nor verified with a listening test. While these recent publications

show encouraging results for BFBR, no comprehensive theoretical formulation or analysis of the approach has been presented so far, and nor has it been compared to the HOA-based binaural reproduction approach. Furthermore, it is not yet clear what are the limitations, or important parameters, affecting this approach.

This work makes an initial contribution towards the development of a theoretical design framework for BFBR. A set of guidelines for selecting the number of beamformers in BFBR is developed, employing a novel measure that leads to an informed selection of the number of beamformers. The consequences of selecting too many beamformers are also studied. First, the theoretical relation between BFBR and HOA-based reproduction is formulated for spherical arrays. Then, an appropriate choice for the number of beamformers is shown to be directly related to the directivity factor (DF) of the maximum directivity (MD) beamformer of spherical arrays. It is then proposed to generalize this measure for general arrays by quantifying the *average directivity factor* of the beam patterns. Next, the use of a larger number of beamformers is discussed and analyzed, showing that it can lead to order truncation of the HRTF and attenuation of high frequencies. Objective analysis using computer simulations was performed based on spherical and planar arrays, and a listening test was conducted to validate the results.

## 2 Overview of BFBR

This section presents the mathematical background necessary for understanding BFBR. The first part of this section introduces a model of the sound pressure measured by general microphone arrays. Then, the BFBR approach is presented for the case of microphone domain beamforming.

Consider a microphone array positioned in a sound field with  $S$  far-field sources. The sound pressure at the  $Q$  microphones is measured at wave number  $k$ , where  $k = \frac{\omega}{c}$ , with  $c$  denoting the speed of sound propagation and with  $\omega$  denoting the radial frequency. The sound pressure vector can be written as

$$\mathbf{p} = \mathbf{V}\mathbf{s} + \mathbf{n}, \quad (1)$$

with

$$\begin{aligned} \mathbf{p} &= [p_1(k) \ p_2(k) \ \cdots \ p_Q(k)]^T \\ \mathbf{s} &= [s(k, \Omega_1) \ s(k, \Omega_2) \ \cdots \ s(k, \Omega_S)]^T \\ \mathbf{n} &= [n_1(k) \ n_2(k) \ \cdots \ n_Q(k)]^T, \end{aligned} \quad (2)$$

and

$$\begin{aligned} \mathbf{V} &= [\mathbf{v}(k, \Omega_1) \ \mathbf{v}(k, \Omega_2) \ \cdots \ \mathbf{v}(k, \Omega_S)] \\ \mathbf{v}(k, \Omega) &= [v_1(k, \Omega) \ v_2(k, \Omega) \ \cdots \ v_Q(k, \Omega)]^T, \end{aligned} \quad (3)$$

where  $n_q(k)$  is the measurement noise of microphone  $q$  at wave number  $k$ .  $s(k, \Omega)$  is the representation of all

far-field source signals positioned at direction  $\Omega$ , with  $v_q(k, \Omega)$  denoting the acoustic transfer function between  $s(k, \Omega)$  and the pressure  $p_q(k)$  measured by microphone  $q$ . The vector  $\mathbf{v}(k, \Omega)$  denotes the acoustic transfer function from source direction  $\Omega$  to the microphone array.  $\mathcal{S} \triangleq \{\Omega_\xi\}_{\xi=1}^S$  is the set of source directions of arrival with the  $\xi^{\text{th}}$  direction given by  $\Omega_\xi = (\theta_\xi, \phi_\xi)$ , where  $\theta_\xi$  denotes the elevation and  $\phi_\xi$  denotes the azimuth direction. The source signals can represent, for example, any combination of direct sound sources and their reflections from room boundaries. Note that throughout this paper, the measurement noise is assumed to be negligible and is therefore omitted for simplicity.

BFBR aims to reproduce binaural signals at the ears from signals captured by a microphone array. This is achieved by employing the spatial information derived when applying beamforming, also denoted as spatial filtering, on microphone array signals [35]. The beamformer output for a steering direction  $\Omega_d$  and wave number  $k$  is computed by

$$y(k, \Omega_d) = \mathbf{w}^H \mathbf{p}, \quad (4)$$

with the beamformer weight vector denoted as

$$\mathbf{w} = [w_1(k, \Omega_d) w_2(k, \Omega_d) \cdots w_Q(k, \Omega_d)]^T. \quad (5)$$

After beamforming is performed at a given set of  $D$  steering directions  $\mathcal{D} \triangleq \{\Omega_d\}_{d=1}^D$ , BFBR is applied for the left and right ears by multiplying each output with the corresponding HRTF, as follows [26]:

$$\hat{p}_{l,r}(k) = \sum_{d=1}^D \alpha_d h_{l,r}(k, \Omega_d) y(k, \Omega_d). \quad (6)$$

Here,  $y(k, \Omega_d)$  is the beamformer output at steering direction  $\Omega_d$ , as presented in (4);  $h_{l,r}(k, \Omega_d)$  denotes the HRTF due to a plane wave arriving from direction  $\Omega_d$  at the left or right ear [16]; and  $\mathcal{A} \triangleq \{\alpha_d\}_{d=1}^D$  denotes the set of weights associated with the set of steering directions  $\mathcal{D}$ , to be discussed later.

The BFBR Eq. (6) provides an approximation for the pressure at the ears, denoted as  $\hat{p}_{l,r}(k)$ , with an accuracy which may depend on the sound field, the set of the steering directions used, the type of beamformers applied, and the characteristics of the microphone array. A comprehensive investigation into the complex, multi-featured behavior and performance of the BFBR approach is sorely needed; this paper takes an initial step towards this by studying the selection of the number of steering directions when capturing complex sound fields with several sources and room reflections. Because such complex sound fields are of interest in this work, sound field formulations based on general representations are employed. This is in contrast to a simpler sound field with only a few known

source directions, for which tailored beamformers can be designed.

### 3 BFBR for spherical microphone arrays

In Section 2, BFBR was introduced for general microphone arrays. In this section, spherical microphone arrays are discussed within the BFBR framework. Spherical microphone arrays are widely used and studied due to their ability to perform decomposition of the sound field into SH, leading directly to the HOA signal. This property is utilized here in order to derive a relation between HOA-based reproduction and BFBR for such arrays. First, BFBR is applied to spherical microphone arrays. Then, conditions for the equivalence of BFBR and HOA-based reproduction are presented.

For a spherical array, the acoustic transfer matrix can be represented in closed form [14, 36]:

$$\mathbf{V} = \mathbf{Y}_Q \mathbf{B} \mathbf{Y}_S^H, \quad (7)$$

with

$$\mathbf{B} = \text{diag}(\mathbf{b}_n)$$

$$\mathbf{Y}_Q = [\mathbf{y}_{nm}(\Omega_1) \mathbf{y}_{nm}(\Omega_2) \cdots \mathbf{y}_{nm}(\Omega_Q)]^T \quad (8)$$

$$\mathbf{y}_{nm}(\Omega) = [Y_0^0(\Omega) Y_1^{-1}(\Omega) \cdots Y_{N_p}^{N_p}(\Omega)]^T,$$

where  $Y_n^m(\Omega)$  denotes the order  $n$  and degree  $m$  SH basis function [14],  $\mathbf{Y}_Q$  is referred to as a SH matrix and is composed of a set of SH vectors  $\mathbf{y}_{nm}(\Omega)$  with the set of microphone directions on the sphere,  $\mathcal{Q} = \{\Omega_q\}_{q=1}^Q$ , and with the maximum SH order  $N_p$  denoting the array order.  $\mathbf{Y}_S$  is similarly defined, where  $\mathcal{S}$  is the set of source directions described in (3).  $\mathbf{b}_n$  is a vector containing the radial functions [14, 37] which are arranged with respect to the SH order of  $\mathbf{y}_{nm}(\Omega)$ . Therefore, for a spherical array,  $\mathbf{V}$  can be represented in terms of the order-limited linear transformation from the SH subspace of the source signal directions to the SH subspace of the microphone directions. The sound pressure model for a spherical microphone array is therefore given by substituting (7) into (1):

$$\mathbf{p} = \mathbf{Y}_Q \mathbf{B} \mathbf{Y}_S^H \mathbf{s} = \mathbf{Y}_Q \mathbf{B} \mathbf{a}_{nm}, \quad (9)$$

with

$$\mathbf{a}_{nm} = [a_{00}(k) a_{1(-1)}(k) \cdots a_{N_p N_p}(k)]^T \quad (10)$$

representing the HOA signal vector, limited by the array's SH order  $N_p$ .

The beamformer output for a spherical microphone array is computed by substituting the sound pressure in Eq. (9) into the beamformer output Eq. (4):

$$y(k, \Omega_d) = \mathbf{w}^H \mathbf{Y}_Q \mathbf{B} \mathbf{a}_{nm}. \quad (11)$$

Next, a beamformer is designed to extract the HOA signal  $\mathbf{a}_{nm}$  (i.e., perform PWD [14]). The MD beamformer has been previously shown to perform PWD of the sound

field surrounding an array [14, 26, 38]. The weights for this MD beamformer can be derived by maximization of the DF, or by using the minimum variance distortionless response (MVDR) beamformer under the assumption of an additive diffuse noise field [39]. For the beamformer output in Eq. (11), the following weights perform PWD:

$$\mathbf{w} = [\mathbf{y}_{\mathbf{nm}}(\Omega_d)^T \cdot (\mathbf{Y}_{\mathcal{Q}}\mathbf{B})^\dagger]^H. \quad (12)$$

Substituting (12) into (11) leads to

$$y(k, \Omega_d) = \mathbf{y}_{\mathbf{nm}}(\Omega_d)^T \cdot \mathbf{a}_{\mathbf{nm}} \equiv \bar{a}(k, \Omega_d), \quad (13)$$

with  $\bar{a}(k, \Omega_d)$  denoting the order limited plane-wave amplitude density function (PWADF) at direction  $\Omega_d$ , which is limited to order  $N_p$  in the SH domain.

BFBR for the spherical microphone array is now applied by substituting Eq. (13) for the set of steering directions  $\mathcal{D}$  into the BFBR Eq. (6):

$$\hat{p}_{l,r}(k) = \sum_{d=1}^D \alpha_d h_{l,r}(k, \Omega_d) \bar{a}(k, \Omega_d). \quad (14)$$

By assuming that the SH representation of the HRTF is limited to order  $N_h$ , and by applying the sampling conditions for aliasing-free integration, as developed in [40], to the set of steering directions  $\mathcal{D}$ , and the set of weights  $\mathcal{A}$  in (14), the BFBR Eq. (14) can be reformulated as an integral:

$$p_{l,r}(k) = \int_{\mathbb{S}^2} h_{l,r}(k, \Omega) \bar{a}(k, \Omega) d\Omega, \quad (15)$$

with  $\mathbb{S}^2$  denoting the surface of the unit sphere. For the reformulation above to hold, the sampling scheme used for the beamformer steering directions must support aliasing-free sampling of order  $N_D = \max\{N_p, N_h\}$  or higher. By applying Parseval's Theorem for SH, followed by the complex conjugate relation [14] to (15) the sound pressure at the ears can be computed using SH representation as [2]

$$p_{l,r}(k) = \sum_{n=0}^N \sum_{m=-n}^n (-1)^m h_{nm}^{l,r}(k) a_{n(-m)}(k), \quad (16)$$

with  $N = \min\{N_p, N_h\}$ . Here,  $h_{nm}^{l,r}(k)$  and the HOA signal  $a_{nm}(k)$  are the spherical Fourier transform (SFT) coefficients of the HRTF and of the PWADF, respectively, employed in (15). This is exactly the sound pressure at the ears, as produced by using the HOA-based reproduction approach [2].

Although this equivalence has only been derived for spherical arrays, or arrays that can perform PWD, this result motivates the derivation of a measure that aids in the selection of the number of beamformers for BFBR with general arrays. Such a measure is suggested in the following section. Furthermore, a known drawback of HOA-based reproduction for HOA signals of low order

is the reduced quality, due to the spatial and temporal smoothing that is a result of the truncation to the HRTF SH order. The latter may greatly reduce the reproduction quality [12, 13]. Section 5 will demonstrate how such smoothing is also likely to occur when employing BFBR with general arrays.

#### 4 Sampling conditions for BFBR with general arrays

Sampling conditions for spherical arrays that guarantee BFBR to be equivalent to HOA-based reproduction have been presented in the previous section. These are outlined here as a starting point for the subsequent discussion:

- (i) PWD beamformers are used.
- (ii) The steering directions and their associated set of weights  $\mathcal{D}$  and  $\mathcal{A}$ , respectively, must define a sampling set on the sphere that is aliasing-free to order  $N_D = \max\{N_p, N_h\}$ , where  $N_p$  is the order of the spherical array and  $N_h$  is the order of the HRTF.
- (iii) Then, BFBR is equivalent to HOA-based reproduction of order  $N = \min\{N_p, N_h\}$ .

However, arrays of general configurations cannot typically perform PWD, and so for such arrays, these sampling conditions are no longer useful. Therefore, we would like to develop a similar set of mathematical conditions for BFBR with general arrays, as such conditions do not exist in the literature. The development of a mathematically complete set of sampling conditions to support BFBR design with general array configurations is challenging and is suggested for future studies. Instead, in this paper, the sampling conditions presented above for spherical arrays are reformulated, and this reformulation forms the basis for the proposal of design guidelines for BFBR with general arrays.

##### 4.1 Reformulation of the sampling conditions for BFBR with spherical arrays

The first step in the reformulation is a simplification with regard to SH orders. Since the order of the array is commonly lower than that of the HRTF, the latter is truncated to equal the order of the array. This reduces the number of steering directions used without changing the reproduced order, i.e.,  $N = \min\{N_p, N_h\}$ . Second, because the SH order is a parameter that is available for spherical arrays, but not for general arrays, we propose substituting the DF of the MD beamformer, which is used for PWD, with the order, as the former can be computed for any array. Now, the DF of the MD beamformer of a spherical array is given by [14]

$$DF(k, \Omega_d) = (N_p + 1)^2, \quad (17)$$

where the DF is generally defined as [35]



$$DF(k, \Omega_d) = 4\pi \frac{|\mathbf{w}^H \mathbf{v}(k, \Omega_d)|^2}{\int_{\mathbb{S}^2} |\mathbf{w}^H \mathbf{v}(k, \Omega)|^2 d\Omega}, \quad (18)$$

with  $\mathbf{v}(k, \Omega_d)$  as in (3), and  $\Omega_d$  is the beamformer steering direction. Using (17) leads to the following, reformulated but equivalent, sampling conditions for spherical arrays:

- (i) MD beamformers are used
- (ii) The steering directions and their associated set of weights  $\mathcal{D}$  and  $\mathcal{A}$ , respectively, each with a cardinality of  $D$ , must define a sampling set on the sphere that leads to aliasing-free integration, satisfying  $D \geq DF$

These conditions are equivalent to the ones presented previously, since the minimum number of samples for integration on the sphere, that is aliasing-free up to order  $N_p$  is  $(N_p + 1)^2$  [41], which is exactly the value of the DF in this case(17). Therefore, also under these conditions, equivalence between BFBR and HOA-based reproduction is obtained.

#### 4.2 Selecting the number of beamformers for general arrays

The reformulated sampling conditions in Section 4.1 form the basis for conditions applied to general arrays. Although they only apply to spherical arrays, they present a relation between the DF and the number of steering directions that lead to good quality reproduction (HOA-based binaural reproduction). Here, we make the following important step: we propose to use the same relations for general arrays. This does not guarantee reproduction in the form of HOA-based reproduction, but could lead to a good design.

However, when BFBR is applied to general arrays, the DF of individual beamformers may depend on the steering direction and on frequency. To overcome this complexity, the *average directivity factor* is proposed as an alternative measure to the DF in the reformulated sampling conditions above and is given by

$$DF_{avg}(k) = \frac{1}{4\pi} \int_{\mathbb{S}^2} DF(k, \Omega) d\Omega. \quad (19)$$

This now leads to sampling conditions that are similar to those presented above for spherical arrays:

- (i) MD beamformers are used
- (ii) The steering directions and their associated set of weights  $\mathcal{D}$  and  $\mathcal{A}$ , respectively, each with a cardinality of  $D$ , must define a sampling set on the sphere that leads to aliasing-free integration, satisfying  $D \geq DF_{avg}(k)$

This choice makes sense for a number of reasons. First, when using a sampling scheme that is directly related to the DF, which is related to the main lobe width, we expect

the set of beams to cover the entire directional space with the appropriate amount of overlap. However, because the DF may change with the look direction, the standard deviation of the DF can be introduced as a useful measure:

$$DF_{STD}(k) = \sqrt{\frac{1}{4\pi} \int_{\mathbb{S}^2} [DF(k, \Omega) - DF_{avg}(k)]^2 d\Omega}. \quad (20)$$

Furthermore, it has been shown in [42, 43] that the average directivity factor of the MD beamformer along all directions is given by the number of microphones in the array, denoted by  $Q$ . This provides a useful simplification — the number of beamformers can be chosen to be equal to or larger than the number of microphones in the array. Therefore, the following even more general relation can replace the sampling conditions for a general array:

- (i) MD beamformers are used
- (ii) The steering directions and their associated set of weights  $\mathcal{D}$  and  $\mathcal{A}$ , respectively, each with a cardinality of  $D$ , must define a sampling set on the sphere that leads to aliasing-free integration, satisfying  $D \geq Q$

The sampling conditions above were formulated by assuming MD beamformers because of the mathematical relations available for these beamformers. This is not to say that these conditions will not hold when using other beam patterns. Other beam patterns can also be applied in practice, but their study is beyond the scope of this paper and is proposed for future work. Furthermore, the average directivity factor may change with frequency. This implies that different frequencies may require different beamformer distributions when using BFBR. Also, frequency-dependent beamformers may lead to timbral distortion, which may affect the performance of BFBR. These issues are also beyond the scope of this paper and are left for future research.

#### 4.3 BFBR with high order HRTFs

The sampling conditions introduced above for BFBR with MD beamformers for general microphone arrays were based on the simplification that a low order HRTF, i.e.,  $N_h = N_p$ , was used. However, the truncation of the order of the HRTF may reduce the spatial qualities, degrading spatial perception [2, 13, 15, 44]. A number of approaches such as Mag-LS [18], spectral equalization [44], or other approaches for spectral coloration correction [12, 19] may be used to improve perception with low order HRTFs. Furthermore, a study has recently compared the above methods [20]. The use of all these methods for BFBR is interesting, but may be beyond the scope of this paper. Here, however, we propose to use the high order HRTF directly without increasing the number of beamformers.

The rationale behind this is that although violating the sampling condition (following [40]), the effect of the high orders on aliasing may be small due to the natural HRTF order roll-off, in particular at the lower frequencies where equivalence to HOA-based reproduction may still approximately hold [2, 40], and the overall effect on perception may be positive [2, 13, 44]. This was verified in an informal listening test.

This alternative is based on two conditions:

- (i) The full high order HRTF is employed.
- (ii) The reformulated sampling conditions in Section 4.1 are used where the number of beamformers  $D$  is chosen to be as close to the lower bound,  $DF_{avg}(k)$ , as possible.

In summary, the use of full-order HRTFs with the minimal number of beamformers aims to strike a balance between keeping high order spatial information and reducing spatial errors due to spatial aliasing. While only partially justified in theory, this approach is shown to be useful, as will become evident from the investigations reported in the following sections.

## 5 The consequences of oversampling with high order HRTFs

In this section, the consequences of oversampling in BFBR are discussed. In the context of this work, oversampling is relative to the conditions in Section 4.3, in conjunction with the reformulated sampling conditions in Section 4.2.

### 5.1 Oversampling leads to order truncation

Assume a beamformer is applied to a general array with an output given by  $y(k, \Omega)$ , as in (6). It is also assumed that the SH representation of  $y(k, \Omega)$  is of limited order denoted by  $N_y$  (see formulation in the next subsection). The HRTF,  $h_{l,r}(k, \Omega)$ , is also assumed to be limited to some high order  $N_h$ , as discussed in Section 4.3. Similar to the case of spherical arrays, a dense beamformer distribution in (6) is assumed such that  $D$  is very large. The latter means that the beamformers' steering directions and their appropriate weights represent a sampling scheme which is aliasing-free up to SH order  $N_D \geq \max\{N_y, N_h\}$ . Under these assumptions, (6) can be replaced by an integral over the sphere, and after applying Parseval's relation, (6) can be rewritten as

$$\hat{p}_{l,r}(k) = \sum_{n=0}^{N_e} \sum_{m=-n}^n (-1)^m h_{nm}^{l,r}(k) y_{n(-m)}(k), \quad (21)$$

where  $N_e = \min\{N_y, N_h\}$  and  $y_{nm}(k)$  are the SFT coefficients of the beamformer output function  $y(k, \Omega)$ .

The similarity between (21) and (16) is clear, where  $y_{nm}(k)$  is replaced by  $a_{nm}(k)$  in Eq. (21). Now, recall that (16) leads to order truncation, as noted above. It is

therefore expected that  $y_{nm}(k)$ , if limited to a low order,  $N_y$ , will lead to a similar effect, i.e., order truncation of the HRTF to order  $N_y$ . This is further analyzed in the next subsection.

### 5.2 Analysis of the order of $y_{nm}(k)$

The function  $y(k, \Omega)$  represents the way in which the beamformer output varies with look direction and therefore may depend on the design of the beamformers. With the aim of gaining insight into the behavior of  $y_{nm}(k)$ , the beamformers are assumed to belong to a common family of designs, including the MVDR (i.e., Capon), and MD, as detailed in Section 3. A general formulation for this beamformer family is given by [35]

$$\mathbf{w}(k, \Omega_d) = c(k, \Omega_d) \cdot [\mathbf{R}(k)]^{-1} \cdot \mathbf{v}(k, \Omega_d), \quad (22)$$

where  $c(k, \Omega_d)$  is a normalization coefficient, and  $\mathbf{R}(k)$  is a noise covariance matrix. For the MVDR beamformer,  $\mathbf{R}(k)$  represents the covariance of the noise signal, while for the MD beamformer, it is assumed that the noise signal is spatially white, or diffuse. Finally, for the Bartlett beamformer,  $\mathbf{R}(k)$  is the identity matrix.

Substituting (22) into (4) and equating the normalization coefficient  $c(k, \Omega_d)$  to be 1 for simplicity lead to

$$y(k, \Omega) = [\mathbf{v}(k, \Omega_d)]^H [\mathbf{R}(k)]^{-1} \mathbf{p}(k) = \sum_{q=1}^Q r_q(k) \cdot [v_q(k, \Omega)]^*, \quad (23)$$

where  $r_q(k)$  are the elements of the product of  $[\mathbf{R}(k)]^{-1}$  and  $\mathbf{p}(k)$ .

It is clear from (23) that the beamformer output in this case is a linear combination of the acoustic transfer functions of the array microphones,  $v_q(k, \Omega)$ . It is therefore expected that  $y_{nm}(k)$  will inherit its SH order behavior from  $v_q(k, \Omega)$ . As discussed in [37], for example, the SH behavior of the acoustic transfer function of a free-field microphone positioned at a distance  $r_a$  from the origin is dominated by radial functions (spherical Bessel and Hankel functions). This behavior typically exhibits a roll-off along the order  $n$ , with a cut-off which is approximately at  $N_p = kr_a$ . However, the latter may depend on whether the microphone is in free-field or placed around a rigid body, for example. On the other hand,  $[\mathbf{R}(k)]^{-1}$  may lead to enhancement of higher orders at the expense of robustness [45], which may lead to noise-sensitive beamformers. Moreover, note that the rank of the matrix  $\mathbf{R}(k)$  is limited by the number of microphones  $Q$ . This further limits the possible order of  $y_{nm}(k)$ .

This analysis leads to the conclusion that the order  $N_y$  of  $y_{nm}(k)$  is therefore expected to be bounded, with the bound affected by the order of the array,  $N_p$ , the type of beamformer, and the number of microphones.

### 5.3 Computation of the order of $y_{nm}(k)$

While the order of  $y_{nm}(k)$  is expected to be limited, as argued in the previous subsection, it may be useful to compute it numerically for specific arrays. A numerical approach for the approximation of  $N_y$  is formulated next.

First,  $y_{nm}(k)$  is computed from the beamformer output function  $y(k, \Omega)$ . After  $y_{nm}(k)$  is obtained, an estimate of  $N_y$  is computed as follows:

$$N_y(k) = \arg \min_N \left\{ \sum_{n=0}^N \sum_{m=-n}^n |y_{nm}(k)|^2 > \eta \cdot \sum_{n=0}^{\infty} \sum_{m=-n}^n |y_{nm}(k)|^2 \right\}, \quad (24)$$

where  $\eta$  is chosen to be of a value close to 1, for which most of the power of the beamformer output is contained up to order  $N_y(k)$ .  $N_y(k)$  can then be used to assess the truncation order of the HRTF when using BFBR.

## 6 Measures of frequency-dependent performance

Truncation of the HRTF, as discussed above, may lead to attenuation at high frequencies. It is therefore desirable to measure the frequency-dependent power of the sound pressure at the ears when applying BFBR, in order to quantify such attenuation. Furthermore, it may also be instructive to compare the accuracy of BFBR relative to a HOA-based reproduction reference. Therefore, the relative mean square error (rMSE) for the BFBR signals is also presented.

To derive the first measure, Eq. (4) is substituted into the BFBR Eq. (6), and its expected power is computed, leading to

$$\begin{aligned} E_{\hat{p}_{l,r}}(k) &= \mathbb{E} \left[ |\hat{p}_{l,r}(k)|^2 \right] = \mathbb{E} \left[ |\mathbf{w}_h^H \mathbf{p}|^2 \right] \\ &= \mathbf{w}_h^H \mathbf{R}_p \mathbf{w}_h, \end{aligned} \quad (25)$$

with  $\mathbf{R}_p$  denoting the covariance matrix of  $\mathbf{p}$ , and

$$\mathbf{w}_h = \sum_{d=1}^D [\alpha_d h_{l,r}(k, \Omega_d)]^* \mathbf{w}. \quad (26)$$

We recall that  $\mathbf{w}$  is dependent on the steering direction  $\Omega_d$  as in (5), and both  $\mathbf{w}$  and  $\mathbf{p}$  are frequency dependent.

This measure can be compared to the case of a high order spherical array as a reference, which is computed by replacing  $\hat{p}_{l,r}(k)$  in (25) with the sound pressure computed using HOA-based reproduction, (16). For simplicity, it is assumed that the sound field is diffuse, in which case the expected power is formulated as a function of the HRTF coefficients:

$$E_{p_{l,r}}(k) = \|\mathbf{h}_{nm}\|_2^2, \quad (27)$$

with

$$\mathbf{h}_{nm} = \left[ h_{00}^{l,r}(k) h_{1(-1)}^{l,r}(k) \cdots h_{N_h N_h}^{l,r}(k) \right]^T \quad (28)$$

representing the HRTF SFT coefficients vector, limited by the HRTF's SH order  $N_h$ . This simplified result is obtained since a diffuse sound field forms a white noise signal in the SH domain [14].

The rMSE for the BFBR signals, relative to a HOA-based reproduction signal (16), is simply given by

$$rMSE(\hat{p}_{l,r}(k)) = \frac{\mathbb{E} \left[ |\hat{p}_{l,r}(k) - p_{l,r}(k)|^2 \right]}{\mathbb{E} \left[ |p_{l,r}(k)|^2 \right]}. \quad (29)$$

When computing the rMSE for a diffuse sound field, the denominator of (29) can be replaced by Eq. (27), where  $N_h$  is set to the SH order  $N$  of the HOA-based reproduction signal. If the sound pressure  $\mathbf{p}$  can be represented in the SH domain, such as in the case of rigid spherical arrays or free field microphone arrays [14], the numerator of Eq. (29) could also be represented in closed form.

Comparing (25) for the case of a diffuse sound field with different values of  $D$  to the expected power reference (27) can be useful for understanding the effect of the choice of the number of beamformers in BFBR on the frequency-dependent magnitude of the binaural signal; here, (29) can be further applied to estimate the signal's accuracy.

## 7 Simulation study with spherical and planar arrays

In this section, a simulation study is presented to illustrate the theory and insights developed in Sections 4 and 5. In particular, spherical and planar arrays are simulated using MATLAB version R2019a, with the planar array serving as an example of a non-spherical array configuration. The selection of the number of beamformers is analyzed, together with other factors, and demonstrates the theoretical results. The listening test in the following section complements the simulation study.

### 7.1 Setup

Two microphone arrays were simulated as follows. A spherical microphone array, consisting of  $Q = 12$  uniformly distributed microphones on the surface of a rigid sphere with a radius of 2.12 cm, based on a t-design of order 2 [46] was used. The array can perform aliasing-free PWD up to 5150 Hz. A planar array, which was a uniform array in free field consisting of  $Q = 9$  microphones on a  $3 \times 3$  grid with a distance of 1.5 cm between the microphones, was used. The array, encapsulated within a sphere with radius 2.12 cm, was positioned on the  $yz$  plane. This microphone configuration allows for an aliasing-free processing bandwidth of about 11,500 Hz, with the sound pressure within the array region accurately represented using a SH expansion of order  $N_p = 5$  in this bandwidth, following the relation  $N_p = kr_a$ , as presented in Section 5.2. Both arrays were centered at the origin and were assumed to include negligible sensor noise. HRTFs from

the Cologne HRTF compilation of the Neumann KU-100 [47] of order  $N_h = 32$  were used in Section 7.4 for the computation of the expected pressure power, presented in Section 6. The analysis in this section was performed over the operating frequency range of the planar array for frequencies between 375 and 11,250 Hz.

### 7.2 Methodology

For the spherical and planar arrays, the sound pressure at each microphone was computed using the SH expansion for the sound pressure on a rigid sphere and in free-field, respectively [14], with  $N_p = 5$ , from which the steering vectors described in (3) were derived. Then, the DF was computed using (18), leading to the computation of its statistics, in accordance with (19) and (20). The MD beamformer for the spherical array was computed using (12), by assuming  $N_p = 2$ , and by adding the minimal diagonal loading necessary for numerical matrix inversion. The MD beamformer for the planar array was computed in practice using the MVDR design framework [39], by assuming a diffuse sound field with unit variance [14], and by further following subsection 5.2, i.e., choosing  $c(k, \Omega_d)$  in (22) to be 1.

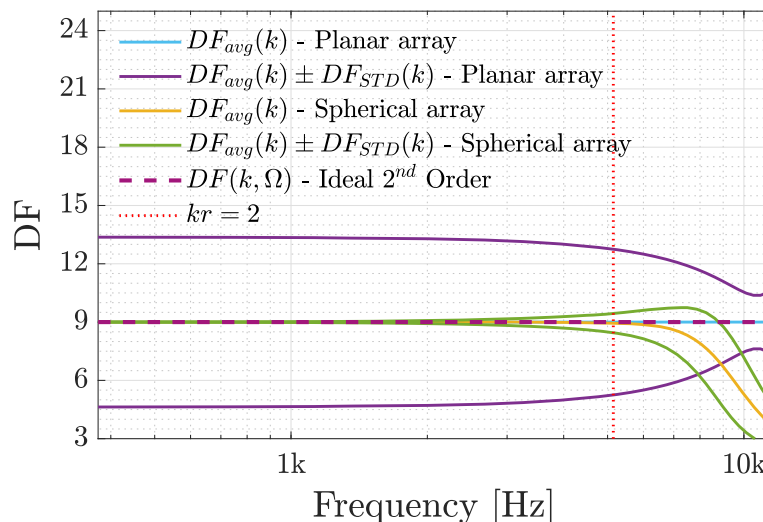
### 7.3 Analysis of DF statistics

In order to illustrate the theory presented in Section 4, the spatial statistics of the DF are presented for the planar and spherical arrays. The average directivity factor (19) and its standard deviation (20) were computed by employing numerical integration on the sphere using a Gaussian sampling scheme [14] of order 60.

Figure 1 shows the average directivity factor and its standard deviation for the two arrays. The figure confirms that for the planar array the average directivity factor

is equal to the number of microphones,  $Q = 9$ , over the entire frequency range, as stated in [42]. The average directivity factor for the spherical array is also 9, as stated in (17), for a limited range; this is because of the significant spatial aliasing above this frequency range when applied to full order sound pressure. Above this frequency range, this beamforming design results in significant degradation to its directivity. Note that by computing the MD beamformers for the spherical array, with the same approach used for the planar array, which is not presented in Fig. 1, the average DF was indeed equal to the number of microphones, i.e., 12. This may imply that a beamformer design, similar to that of the planar array, may improve and extend the operating frequency range of the array. However, this is not further researched in this paper and is suggested for future study. The standard deviation for the planar array changes from  $\pm 4.5$  to  $\pm 1.5$  throughout the frequency range, while for the spherical array the standard deviation is zero, as expected from its constant DF (17) up until around the aliasing frequency of the array. The relatively large standard deviation of the planar array stems from its geometry, where the DF is higher for steering directions in the  $yz$  plane, the array end-fire direction, which is the plane of the array microphones. The reduced variability for the planar array at high frequencies is expected, since the distance between microphones approaches half a wavelength [35]. These computed values for the average directivity factor formed the basis for selecting the number of beamformers in the listening test, as suggested in Section 4.3, in conjunction with the sampling conditions in Section 4.2.

While the average directivity factor is used in this work for the selection of the number of beamformers, as described in Section 4, the standard deviation of the DF



**Fig. 1** The average directivity factor (19) and its standard deviation (20) for the planar and spherical arrays, as described in Section 7.1



may also affect the selection process. However, further investigation of this is out of the scope of this paper and is left for future study.

### 7.4 The effective order of $y_{nm}(k)$

In the previous subsection, both arrays were shown to have an average directivity factor of 9. Following the discussion in Section 5, this means that using a much higher number of beamformers may lead to order truncation of the HRTFs, due to the order limitation of  $y_{nm}(k)$ . In this section, the effective order of  $y_{nm}(k)$  is investigated in order to quantify and gain insight into the effect of order-truncation.

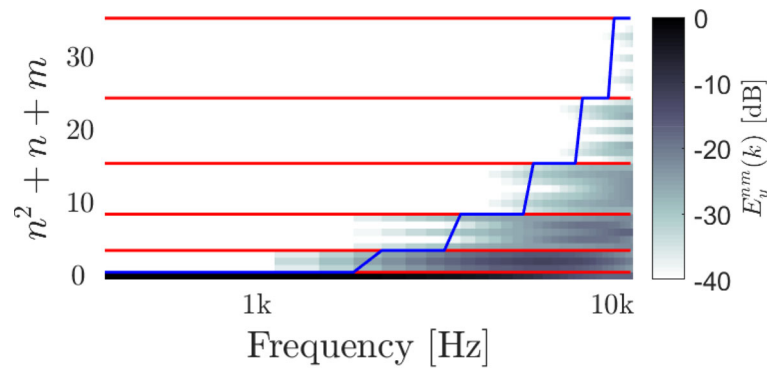
Recall that Eq. (21) in Section 5 presented  $y_{nm}(k)$  as a modified plane-wave density function, taking the role of  $a_{nm}(k)$  in (16). As discussed in Section 5.2, the behavior of  $y_{nm}(k)$  may depend on the beamformer type. Therefore, in this study, MD vs. Bartlett beamformers are evaluated. The effective order of  $y_{nm}(k)$  is computed in practice using (24), as described in Section 5.3.

The effective order of  $y_{nm}(k)$  was computed for the planar array described in Section 7.1, under the assumption

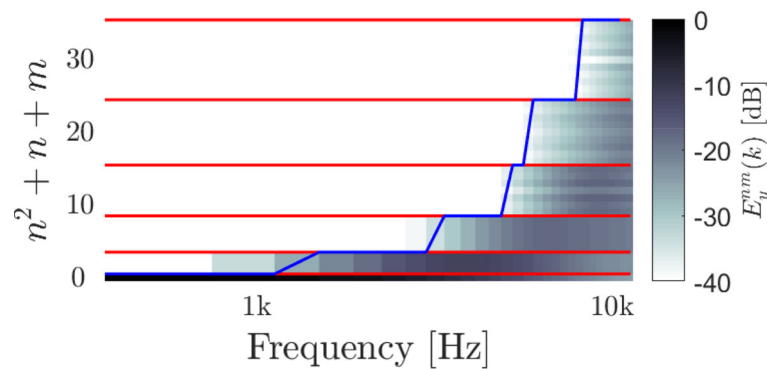
of a diffuse sound field for the two beamformers, and is presented in Figs. 2 and 3. The red lines, located at  $(n+1)^2$ , differentiate between values of  $(n, m)$  having the same order  $n$ . The blue line demonstrates the effective order,  $N_y(k)$ , at each frequency, presented as  $(N_y(k) + 1)^2$ .  $N_y(k)$  was computed using (24) with  $\eta$  chosen to be 0.99. Figures 2a and 3a show the results for the planar array, while Figs. 2b and 3b show the results for the spherical array.

Figure 2a shows the expected power,  $E_y^{nm}(k)$ , for the Bartlett beamformer, with its effective order,  $N_y(k)$ , growing with frequency up to  $N_y(k) = 5$ , as expected from the SH expansion of microphones in free-field. The very low order up to around 2000 Hz is a result of the beamformer's robustness to sensor noise and the behavior of the SFT coefficients of the microphones' acoustic transfer functions [45].

The spherical array, as presented in Section 7.1, is also presented in Fig. 2b for contrast with the planar array in Fig. 2a. Here, also, Bartlett beamformers were used. The behavior of the spherical array is similar to that of the planar array, as expected. Here, however, the first order is attained at a lower frequency. This is because the micro-

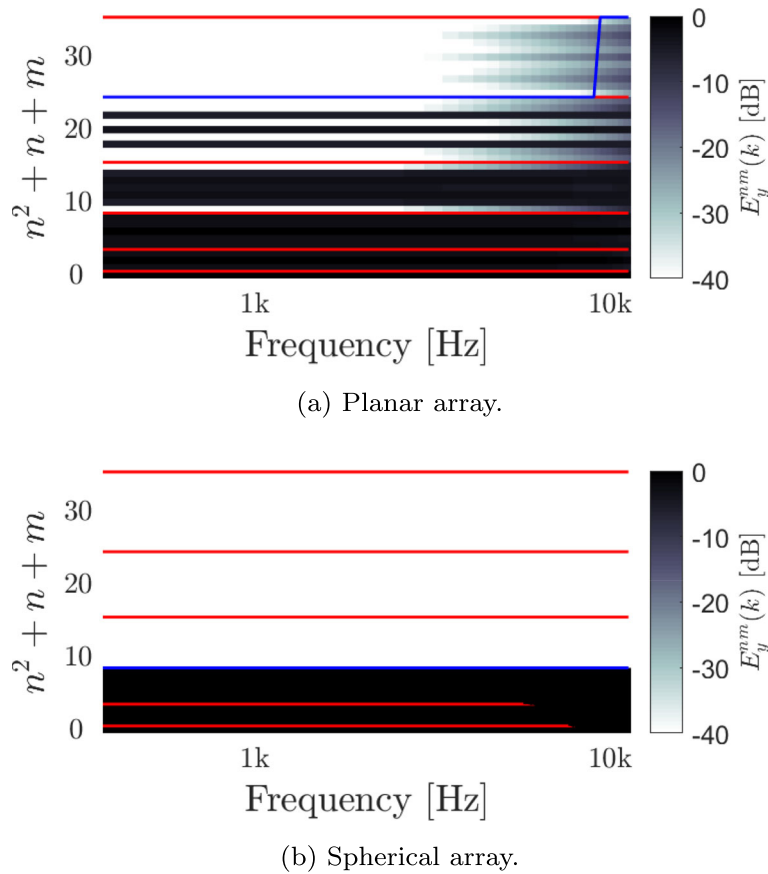


(a) Planar array.



(b) Spherical array.

**Fig. 2**  $E_y^{nm}(k)$ , the expected power distribution of  $y_{nm}(k)$  for the planar and spherical arrays with a diffuse sound field for a Bartlett beamforming design. The red lines, located at  $(n + 1)^2$ , differentiate between different orders. The blue line indicates the order  $N_y(k)$ , presented as  $(N_y(k) + 1)^2$ , and computed as in (24) with  $\eta = 0.99$



**Fig. 3**  $E_y^{nm}(k)$ , the expected power distribution of  $y_{nm}(k)$  for the planar and spherical arrays with a diffuse sound field for a MD beamforming design. The red lines, located at  $(n + 1)^2$ , differentiate between different orders. The blue line indicates the order  $N_y(k)$ , presented as  $(N_y(k) + 1)^2$ , and computed as in (24) with  $\eta = 0.99$

phones of the spherical array have the same radius, which lead to a faster increase of attainable HOA order with frequency.

Figure 3a shows  $E_y^{nm}(k)$  for the MD beamformer. In this case,  $N_y(k)$  was computed to be 4 up to around 9000 Hz, followed by  $N_y(k) = 5$  at higher frequencies. This behavior at the higher orders compared to the Bartlett beamformer can be explained by the dependency of the beamformer weights on the inverse of the covariance matrix  $\mathbf{R}(k)$ . This amplification of high orders at low frequencies leads to a high sensitivity to sensor noise, which is assumed to be negligible in this paper. Furthermore, note that, as discussed in Section 5.2, the maximal rank of the matrix  $\mathbf{R}(k)$  is 9 in this case, therefore limiting the number of independent elements in  $y_{nm}(k)$  to the same number. This, in turn, may impact perception. Similarly, a MD beamformer design with the spherical array is presented in Fig. 3b. As expected, the beamformer output is limited to order 2, which results from its beamforming design approach.

Since both beamformer designs were found to be order limited, it is expected that the use of too many beamformers in BFBR will lead to attenuation at higher frequencies due to the truncation of the HRTF order, as described in Section 5.1.

### 7.5 Spectral analysis of BFBR with different numbers of beamformers

Having illustrated the order limitation of  $y_{nm}(k)$ , and having evaluated the average directivity factor to be 9, it is expected from the theory developed in this paper that HRTF order truncation will occur when selecting a number of beamformers that is significantly higher than 9. This is illustrated in this section.

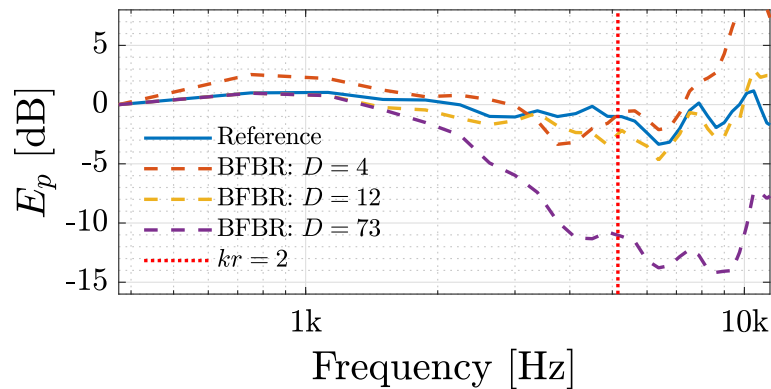
Three sets of beamformers were selected, i.e., 4, representing too few beamformers; 12, representing a number which is approximately the same as the average directivity factor; and 73, which is much higher. The beamforming directions were distributed uniformly, or nearly uniformly, according to appropriate t-designs, facilitating accurate

integration on the sphere up to SH orders of  $N_D = 1, 2,$  and  $5,$  respectively [46]. For the spherical and planar arrays, the expected power of the sound pressure at the ears was computed using (25) and was normalized with the value at frequency 375 Hz. The expected power for HOA-based reproduction signals of order 32 served as a reference and was computed using (27). Both were computed for the left ear and for both arrays and are presented in Fig. 4. The figure shows that for both arrays, increasing the number of beamformers beyond the average directivity factor ( $D = 73$  in the figure) leads to significant attenuation of high frequencies, as expected. Note that this roll-off at high frequencies is significant even though the condition  $N_D \geq \max\{N_y, N_r\}$ , discussed in Section 5.1, did not hold. Furthermore, using a number of beamformers which does not satisfy the conditions in Section 4, i.e.,  $D = 4$  (smaller than 9), is shown to lead to significant amplification at high frequencies. Interestingly, BFBR with 12 beamformers is the most similar to the HOA-based reproduction reference. This implies that using a number

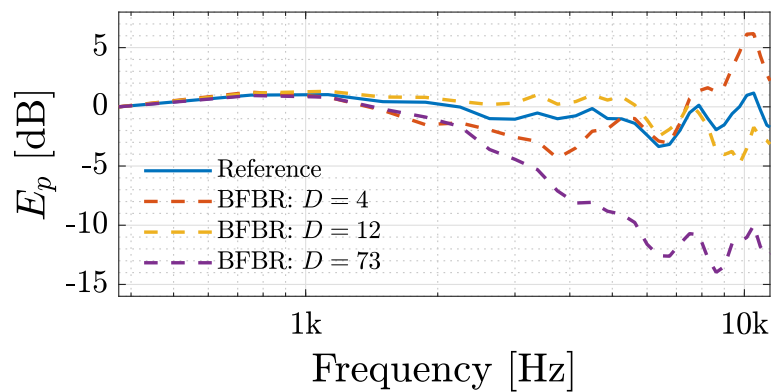
of beamformers that is similar to the average DE, but not much larger, may lead to correct timbre in the reproduced signal. This latter attribute will be further demonstrated in the listening test in Section 8.

To complement the analysis above, the  $rMSE$  of the BFBR signals was also computed, by applying Eq. (29) on both arrays, relative to the HOA-based reproduction reference presented previously in this section, and to 2nd order Ambisonics-based reproduction, i.e., the order of the spherical array, and is presented in Fig. 5. Here, BFBR with the same number of beamformers as presented above is used, with the addition of a BFBR signal with  $D = 2178$  beamformers, distributed using a Gaussian sampling scheme [14] of order 32 — the order of the HRTE, which leads to (21).

The figure shows that for both arrays the lower frequencies have a low  $rMSE$  when the sampling conditions discussed in Section 4 hold. This is expected, as was also discussed in Section 4 and in [40]. Interestingly, the error of the planar array at low frequencies is also relatively low,

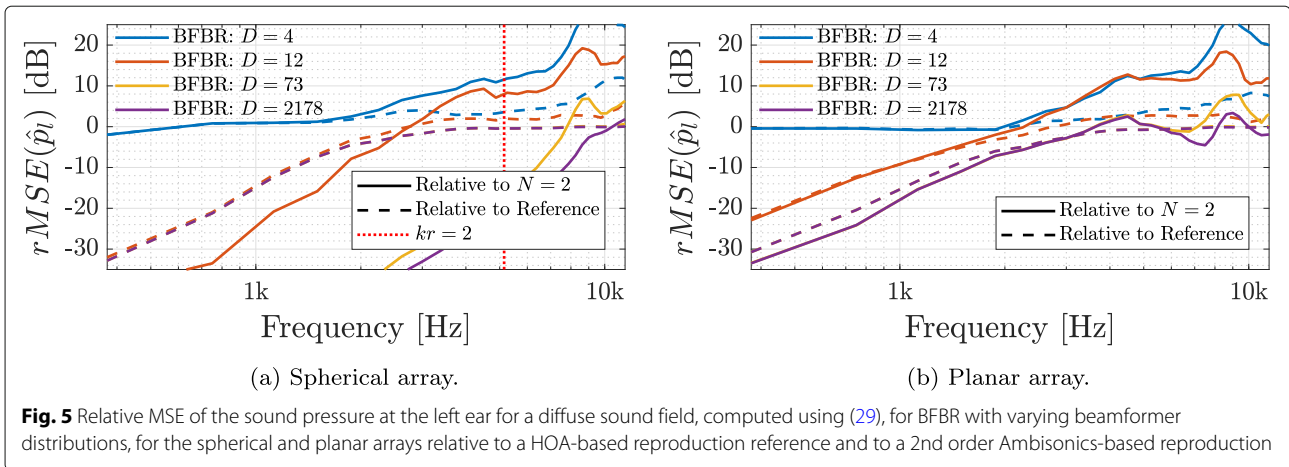


(a) Spherical array.



(b) Planar array.

**Fig. 4** Expected power of the sound pressure at the left ear for a diffuse sound field, computed using (25), for BFBR with varying beamformer distributions, for the spherical and planar arrays. These are compared to a HOA-based reproduction reference of order 32, computed using (27). The signals were further normalized by their value at frequency 375 Hz



even though the array has front-back ambiguity. This may be related to the isotropic properties of the diffuse noise, and so the error may increase in other sound fields. The error relative to the HOA-based reproduction reference reaches saturation at higher frequencies, at around 0 dB for both arrays, while the error relative to the 2nd order Ambisonics-based reproduction becomes much higher at high frequencies. This is expected due to the missing higher orders in the 2nd order Ambisonics-based reproduction signal. These results, along with the previous results in this section, validate the choice of the number of beamformers in BFBR, as suggested in Section 4.

## 8 Listening test

Two listening tests were performed to validate the theory derived in Sections 4 and 5 and to further evaluate spatial perception with BFBR. The tests were based on Recommendation ITU-R BS.1534-1 (MUSHRA, Multiple Stimuli with Hidden Reference and Anchor) [48] and aimed to investigate the effects of the number of beamformers in BFBR on spatial and timbre perception. Due to the current circumstances of COVID-19, the listening test was performed virtually, with each participant performing the test at the place of their choice and with their own equipment.

### 8.1 Setup

The simulated acoustic environment in this test was composed of a rectangular room of dimensions  $9.8 \times 15.5 \times 7.5$  m with a wall reflection coefficient of 0.8 and a reverberation time  $T_{60} = 0.75$  s, resulting in a critical distance of  $r_d = 2.2$  m. The sound field in the room was simulated using the image method [49], where a sampling rate of  $f_s = 48$  kHz was employed.

A planar array was simulated in this room, along with a spherical array of order 2, as described in Section 7.1. The spherical array's origin of coordinates was located 1 m from the wall, at a height of 1.5 m from the floor, at posi-

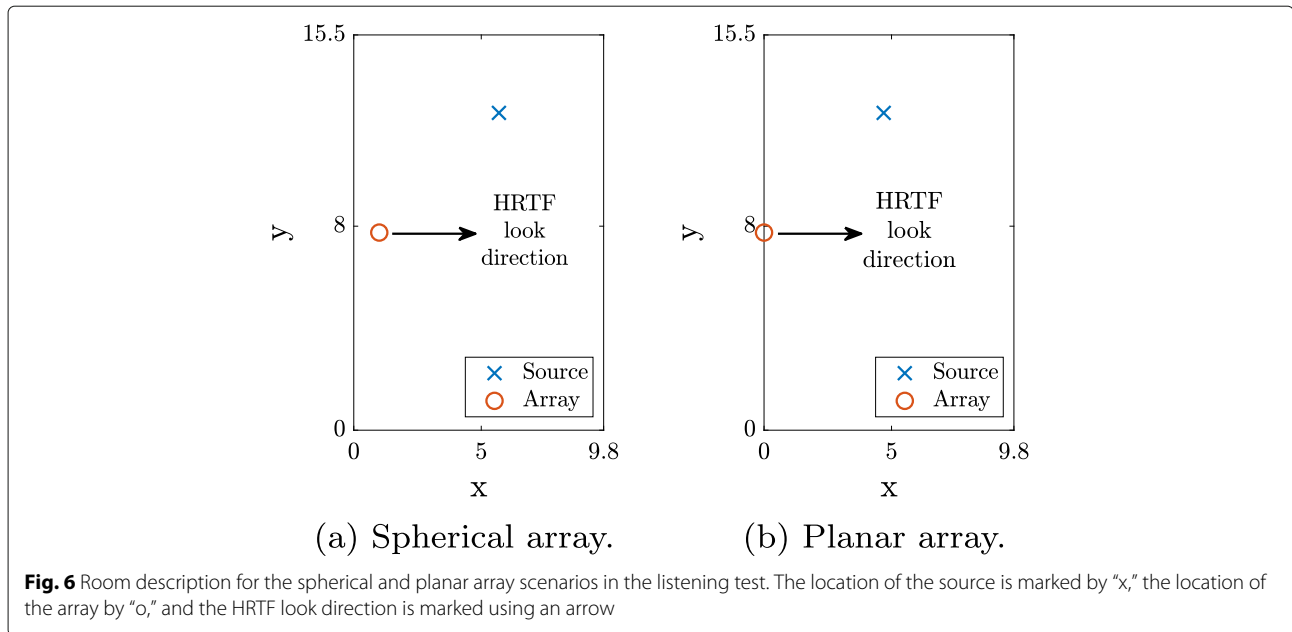
tion (1, 7.75, 1.5) m, while the planar array was positioned on the wall at a height of 1.5 m from the floor, with the array center at (0, 7.75, 1.5) m. The array was positioned on the wall in order to avoid the ambiguity inherent in the directivity of this array. A single source was located at a distance of  $d = 3r_d$  from each array, at an azimuth and elevation of  $(\theta, \phi) = (\frac{\pi}{2}, \frac{\pi}{4})$  relative to the array, i.e., at the same height and at  $45^\circ$  horizontally. Two types of source signals were used. The first signal was speech, taken from the McGill TSP speech database [50], and the second was composed of repeating pulses of noise, each pulse with a duration of 0.2 s, spaced by 0.33 s of silence.

The HRTF used in the tests was the same as that presented in Section 7.1, which was further equalized with a headphone compensation filter [47], selected according to the specific headphones of each user. The signals in the evaluation were played back directly from MATLAB using a set of three different headphones available to the participants, namely, AKG-K701/K702 and Beyerdynamic-DT990PRO. The source and array locations and the HRTF look direction are shown in Fig. 6.

### 8.2 Methodology

The signals for the listening test were generated as follows. Room impulse responses (RIR) were computed for each microphone by first computing the HOA signal's impulse response for each array's origin of coordinates, which was then used to compute the sound pressure of the microphones using plane-wave composition representation [14], assuming free-field for the planar array. For the spherical array, the pressure was computed as in (9) with the HOA signal order limited to 2 to avoid the effect of aliasing. This simplification was introduced to ensure that the comparison between the arrays is focused on the beamforming-based reproduction aspect, rather than other array design aspects. The spherical array is compared to the planar array, as discussed in Section 7.1. Then, MD beamformers were applied to the microphone





signals (the RIRs), as presented in Section 7.2, with the beamformers steered at various directions, as specified below. Finally, the beamformer outputs were filtered by the appropriate HRTFs, as described in (6), generating binaural room impulse responses (BRIRs). These were then convolved with anechoic source signals to produce the binaural signals. The BRIRs were computed at a single head rotation, i.e., the head is directed at  $(\frac{\pi}{2}, 0)$ . HOA-based BRIRs were produced by incorporating the RIR of the source signal as a HOA signal, previously computed for the microphone arrays, into (16), using the same HRTF, later to be convolved with the same anechoic source signals. All the signals generated for the scenarios were further filtered with a low pass filter with a cutoff frequency of 10,865 Hz, in order to avoid spatial aliasing effects on the reproduced signals of the planar array, and for conformity.

Five types of binaural signals were generated:

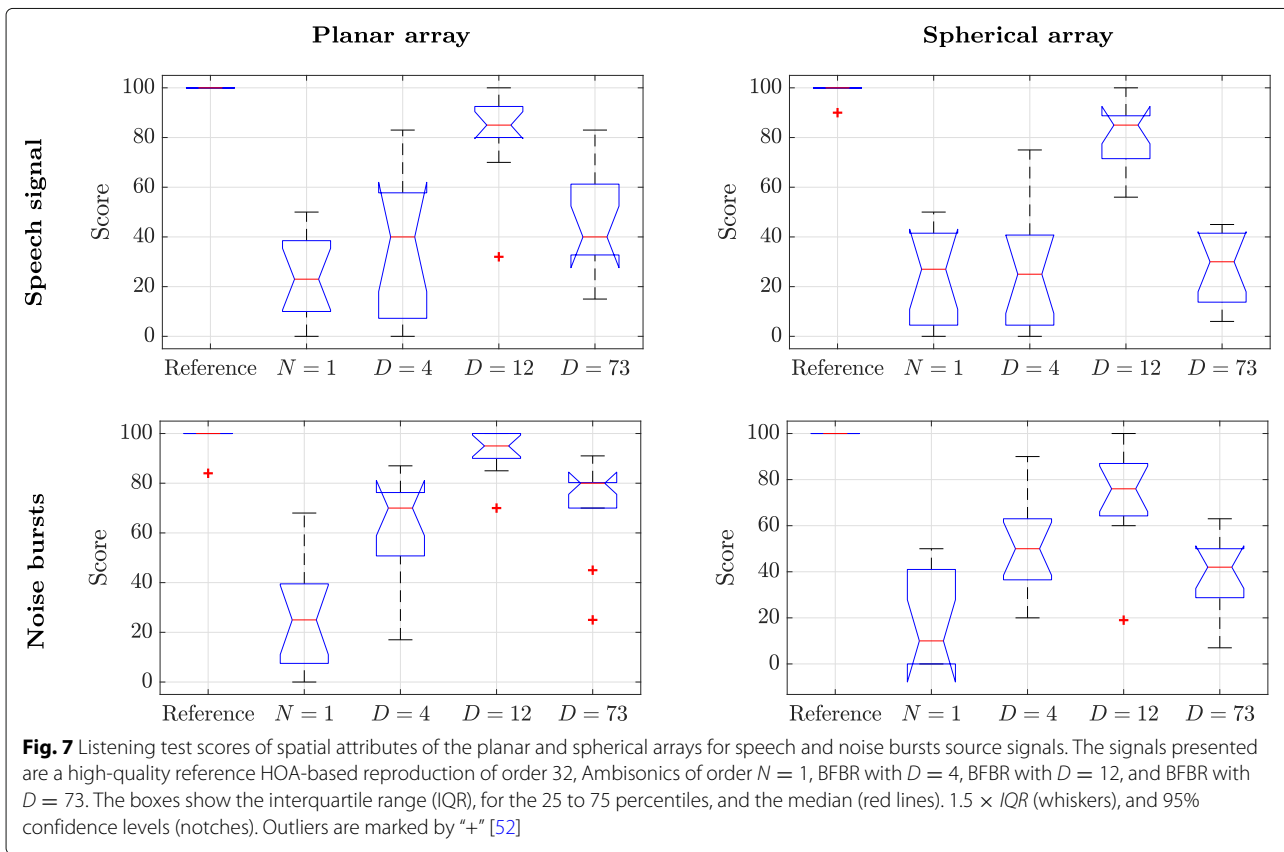
- (i) **Reference:** HOA-based reproduction of order 32
- (ii) **Low order reference:** First order Ambisonics-based binaural reproduction ( $N = 1$ )
- (iii) **BFBR with a sparse beamformer distribution:** BFBR with  $D = 4$  directions, distributed using a uniform sampling scheme of order  $N_D = 1$
- (iv) **BFBR with the recommended number of beamformers:** BFBR with  $D = 12$  directions, distributed using a uniform sampling scheme of order  $N_D = 2$
- (v) **BFBR with a dense beamformer distribution:** BFBR with  $D = 73$  directions, distributed using a nearly uniform sampling scheme of order  $N_D = 5$

The sampling schemes used for signals (iii), (iv), and (v) were previously detailed in Section 7.5 and were applied to both arrays. It should be noted that no anchor signals were used in this experiment. This is for two reasons. First, no obvious anchor was available for these tests. Second, in order to avoid the case where an extreme anchor would reduce relative differences between conditions in the test. Therefore, signal (ii) was also added in order to better illustrate the results in this section.

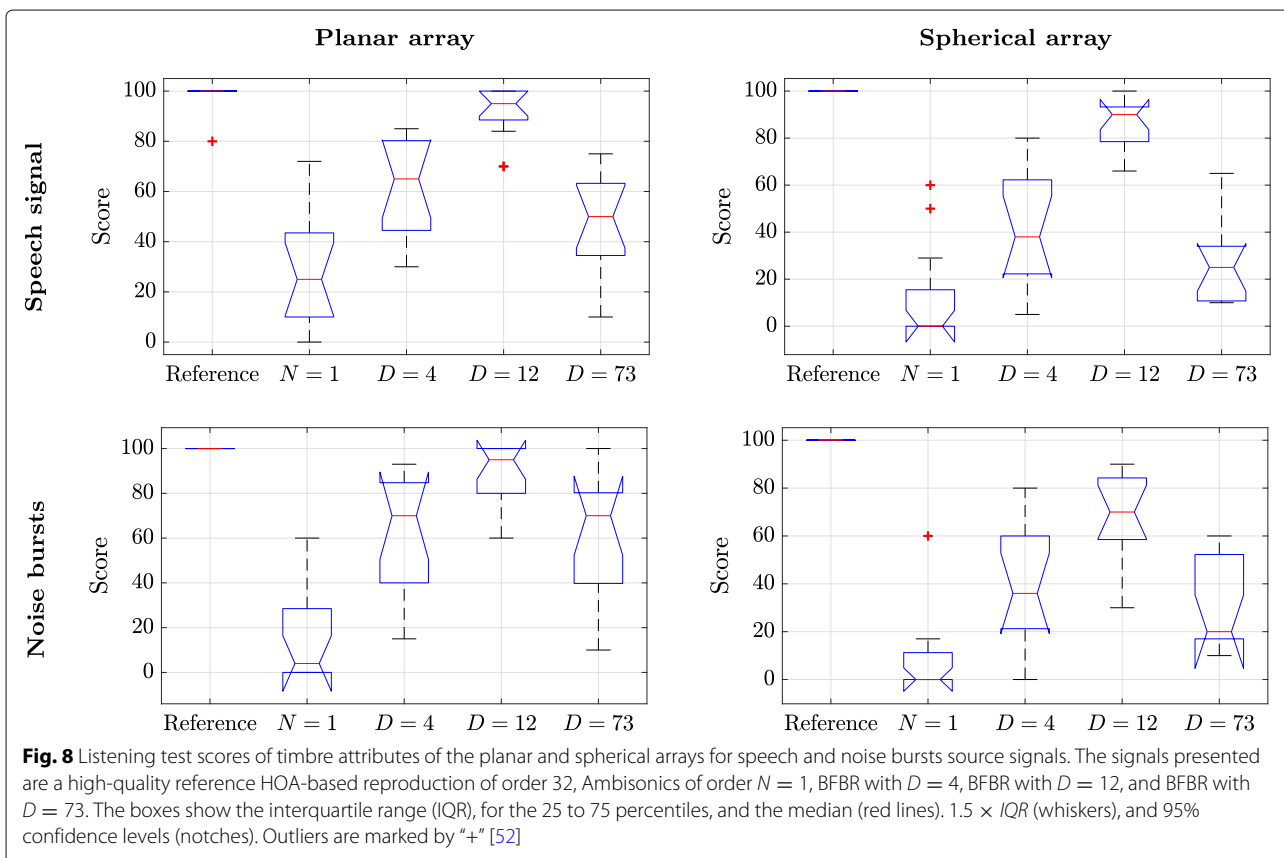
Thirteen subjects with no known hearing impairments participated in the listening tests, where 6 of the participants reported to be expert listeners, 5 reported intermediate expertise, and the remaining two were naive listeners. The listening test comprised 8 screens, each including the 5 signals detailed above. In the first 4 screens, which included all four combinations of the two arrays and the two source signals, the subjects were instructed to score the similarity to the reference with respect to spatial attributes. These included a consolidation of externalization, envelopment, horizontal and vertical direction of the source, and spatial disintegration. The following 4 screens were the same, but here, the subjects were instructed to score the similarity to the reference based on timbre perception, or more specifically, the attribute of tone color. All of these attributes were chosen from [51], with each set of 4 screens presented in a random order.

### 8.3 Results

The results for the spatial attributes and timbre are presented in Figs. 7 and 8, respectively. Both figures show that the hidden reference signal, i.e., signal (i), was correctly identified by the subjects in all screens, except for a few



**Fig. 7** Listening test scores of spatial attributes of the planar and spherical arrays for speech and noise bursts source signals. The signals presented are a high-quality reference HOA-based reproduction of order 32, Ambisonics of order  $N = 1$ , BFBR with  $D = 4$ , BFBR with  $D = 12$ , and BFBR with  $D = 73$ . The boxes show the interquartile range (IQR), for the 25 to 75 percentiles, and the median (red lines).  $1.5 \times IQR$  (whiskers), and 95% confidence levels (notches). Outliers are marked by “+” [52]



**Fig. 8** Listening test scores of timbre attributes of the planar and spherical arrays for speech and noise bursts source signals. The signals presented are a high-quality reference HOA-based reproduction of order 32, Ambisonics of order  $N = 1$ , BFBR with  $D = 4$ , BFBR with  $D = 12$ , and BFBR with  $D = 73$ . The boxes show the interquartile range (IQR), for the 25 to 75 percentiles, and the median (red lines).  $1.5 \times IQR$  (whiskers), and 95% confidence levels (notches). Outliers are marked by “+” [52]

outliers. Furthermore, the figures show that the  $N = 1$  signals, i.e., signal (ii), received the lowest scores overall in both tests. Although this signal was not considered to be an anchor in these listening tests, the above result was rather expected since a low SH order is known to limit the perceived spaciousness and to have an effect on the perceived timbre [13]. The BFBR signals show similar trends within each figure. These are discussed below.

A repeated measures analysis of variance (ANOVA) test [53] was performed for both tests. The majority of variables passed the Kolmogorov-Smirnov normality test [53] except for the reference and a few other conditions which failed due to a ceiling effect.

### 8.3.1 Spatial perception

For the screens concerning spatial perception, signal (iv) (BFBR with  $D = 12$ ), near the recommended number of directions, is the most similar to the reference in all four screens. The difference in scores between signals (iv) and signals (iii) and (v) (BFBR with  $D = 4$  and  $D = 73$ , respectively) is significant, as indicated by the box plots. This is true for both arrays and both source signals and demonstrates that the average directivity factor ( $D = 9$  in this case) is a good measure for the recommended number of beamformers, as suggested in Section 4. The degradation in performance for signal (v) ( $D = 73$ ) compared to signal (iv) ( $D = 12$ ) is a demonstration of the consequence of choosing too many beamformers, as discussed in Section 5. As discussed, the latter leads to order truncation (see Section 7.4), with impact on spatial perception. As the score for the case of  $D = 73$  is higher than the score for  $N = 1$ , the order truncation seems to lead to a SH order higher than 1; as is indeed illustrated in Section 7.4, the order of the planar array SH elements of up to order 4 could be identified, where the spherical array is, of course, limited to order 2. This difference between the planar and the spherical arrays may further explain the differences in scores for signal (v) between the two arrays. The scores for signal (iii) with  $D = 4$  are also significantly lower than for the signals with  $D = 12$ . This is also in line with the theoretical and simulation results presented above. In particular, the use of a too low number of beamformers that only partially cover the entire directional space leads to a noticeable difference from the reference.

A repeated measures ANOVA test, performed for this listening test, resulted in a main effect of the algorithm type independent variable of  $[F(4, 48) = 124, p < .001]$ . The anechoic source signal type (speech vs noise bursts) independent variable showed a main effect of  $[F(1, 12) = 61.9, p < .001]$ , and the array type (planar array vs spherical array) independent variable showed a main effect of  $[F(1, 12) = 27.2, p < .001]$ . A sphericity test leads to a correction only for the algorithm type independent variable. Applying the Greenhouse-Geisser correction to the algo-

rithm type independent variable leads to  $[F(2.52, 30.3) = 124, p < .001, \hat{\epsilon} = .63]$ . However, it can be seen from the results that these corrections did not impact significance levels. Figure 7 shows that the trends for the two source signal types are overall similar. The differences between the two array types may be explained by the better performance of signals (iii) and (v) for the planar array compared to that of the spherical array. While the former showed elements up to order 4, the latter has all elements up to order 2 (see Section 7.4).

The interaction between the source signal type and the array type independent variables showed a main effect of  $[F(1, 12) = 8.54, p = .013]$ , with no necessary correction needed. The interaction between the algorithm type and the source signal type showed a main effect of  $[F(4, 48) = 8.02, p < .001]$ . Applying the Greenhouse-Geisser correction to the latter interaction leads to  $[F(1.57, 18.9) = 8.02, p = .005, \hat{\epsilon} = .39]$ . Similarly, the interaction between the algorithm type and the array type showed a main effect of  $[F(4, 48) = 6.02, p = .001]$ . Applying the Greenhouse-Geisser correction to the interaction above resulted in a corrected main effect of  $[F(2.20, 26.4) = 6.02, p = .006, \hat{\epsilon} = .55]$ . The interaction between all independent variables showed a main effect of  $[F(4, 48) = 1.01, p = .41]$ . Applying the Greenhouse-Geisser correction here leads to a corrected main effect  $[F(2.09, 25.1) = 6.02, p = .38, \hat{\epsilon} = .52]$ . This analysis shows, in particular, that the way in which the binaural reproduction algorithm affects spatial perception may change with array type and signal types, although the trends, as presented in Fig. 7, are similar.

### 8.3.2 Timbre perception

Here, as well, signal (iv) ( $D = 12$ ) was scored significantly higher compared to the other signals. Both signal (iii) ( $D = 4$ ) and signal (v) ( $D = 73$ ) showed a lower score, which demonstrates the theory presented in Sections 4 and 5. Signal (v) ( $D = 73$ ) demonstrates the timbre effects due to the use of too many beamformers, while signal (iii) seems to also suffer from timbre distortion due to the use of a too low number of beamformers. The results validate the simulation results presented in Section 7.5.

A repeated measures ANOVA test was also conducted for timbre. Here, this resulted in a main effect for the binaural reproduction algorithm type independent variable of  $[F(4, 48) = 98.4, p < .001]$ . The anechoic source signal type independent variable showed a main effect of  $[F(1, 12) = 2.90, p = .11]$ , and the array type independent variable showed a main effect of  $[F(1, 12) = 19.5, p = .001]$ . A sphericity test leads to a correction only for the algorithm type independent variable. Applying the Greenhouse-Geisser correction to the latter lead to  $[F(2.77, 33.2) = 98.4, p < .001, \hat{\epsilon} = .69]$ . The above corrections did not impact significance levels.

The interaction between the source signal type and the array type independent variables showed a non-significant main effect of  $[F(1, 12) = 2.68, p = .13]$ , with no necessary correction needed. The interaction between the algorithm type and the source signal type showed a main effect of  $[F(4, 48) = 7.32, p < .001]$ , with a corrected main effect of  $[F(2.65, 31.8) = 7.32, p = .001, \hat{\epsilon} = .66]$ . Similarly, the interaction between the algorithm type and the array type showed a main effect of  $[F(4, 48) = 6.24, p < .001]$ , with the corrected main effect  $[F(3.01, 36.1) = 6.24, p = .002, \hat{\epsilon} = .75]$ . The interaction between all independent variables showed a main effect of  $[F(4, 48) = 2.60, p = .047]$ , with the corrected main effect  $[F(2.96, 35.6) = 2.60, p = .068, \hat{\epsilon} = .74]$ . Similarly to the spatial perception listening test, the analysis above also shows that the effects of the binaural reproduction algorithm on timbre may change with array type and with signal types, although the trends, as presented in Fig. 8, are similar here as well.

## 9 Conclusions

In this study, a novel measure for the recommended number of beamformers in BFBR with general arrays was introduced, and the consequences of beamformer direction oversampling with high order HRTFs were investigated theoretically and analyzed numerically. It was demonstrated, both objectively and perceptually, that beamformer direction oversampling leads to spatial degradation and to the attenuation of high frequencies. Furthermore, objective analysis and listening tests showed that by using a number of beamformers near to that recommended herein, the best binaural reproduction is achieved relative to a high-quality reference. Although this study developed an important beamforming measure for BFBR, the conditions under which it is applicable were only evaluated for a single non-spherical array, and were not investigated for a wide range of arrays, beamforming designs, or acoustic scenarios. Furthermore, a comprehensive investigation on the perceptual differences between such arrays, and an investigation of the effects of their DF patterns is also required. Therefore, such studies are proposed for future work. Furthermore, the development of a more comprehensive design methodology for beam pattern design and look direction selection is also proposed for future work.

### Abbreviations

ANOVA: Analysis of variance; BFBR: Beamforming-based binaural reproduction; BRIR: Binaural room impulse response; DF: Directivity factor; HOA: Higher order Ambisonics; HRTF: Head-related transfer function; IQR: Interquartile range; MD: Maximum directivity; MUSHRA: Multiple Stimuli with Hidden Reference and Anchor; MVDR: Minimum variance distortionless response; PWADF: Plane-wave amplitude density function; PWD: Plane-wave decomposition; RIR: Room impulse response; SFT: Spherical Fourier transform; SH: Spherical harmonics

### Acknowledgements

Not applicable.

### Authors' contributions

II performed the entire research and wrote the paper. BR supervised the research. The authors read and approved the final manuscript.

### Funding

This research was supported in part by Reality Labs Research at Meta and Ben-Gurion University of the Negev.

### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

Received: 14 June 2021 Accepted: 14 February 2022

Published online: 30 March 2022

## References

1. J. Herre, J. Hilpert, A. Kuntz, J. Plogsties, MPEG-H audio—the new standard for universal spatial/3D audio coding. *J Audio Eng Soc.* **62**(12), 821–30 (2015)
2. B. Rafaely, A. Avni, Interaural cross correlation in a sound field represented by spherical harmonics. *J Acoust Soc Am.* **127**(2), 823–28 (2010). <https://doi.org/10.1121/1.3278605>. <https://doi.org/10.1121/1.3278605>
3. M.A. Poletti, Three-dimensional surround sound systems based on spherical harmonics. *J Audio Eng Soc.* **53**(11), 1004–25 (2005)
4. Z. Li, R. Duraiswami, E. Grassi, L.S. Davis, in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays (IEEE, 2004)
5. I. Balmages, B. Rafaely, Open-sphere designs for spherical microphone arrays. *IEEE Trans Audio, Speech, Lang Process.* **15**(2), 727–32 (2007)
6. F. Melchior, O. Thiergart, G. Del Galdo, D. de Vries, S. Brix, in *Audio Engineering Society Convention 127. Dual radius spherical cardioid microphone arrays for binaural auralization* (Audio Engineering Society, 2009)
7. B. Rafaely, The spherical-shell microphone array. *IEEE Trans Audio, Speech, Lang Process.* **16**(4), 740–47 (2008)
8. T.D. Abhayapala, D.B. Ward, in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2. Theory and design of high order sound field microphones using spherical microphone array (IEEE, 2002), p. 1949
9. B. Rafaely, B. Weiss, E. Bachmat, Spatial aliasing in spherical microphone arrays. *IEEE Trans Signal Proc.* **55**(3), 1003–10 (2007)
10. B. Rafaely, Analysis and design of spherical microphone arrays. *IEEE Trans Speech Audio Process.* **13**(1), 135–43 (2004)
11. B. Rafaely, Plane-wave decomposition of the sound field on a sphere by spherical convolution. *J Acoust Soc Am.* **116**(4), 2149–57 (2004)
12. B. Bernschütz, A.V. Giner, C. Pörschmann, J. Arend, Binaural reproduction of plane waves with reduced modal order. *Acta Acustica U Acustica.* **100**(5), 972–83 (2014)
13. A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, B. Rafaely, Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *J Acoust Soc Am.* **133**(5), 2711–21 (2013). <https://doi.org/10.1121/1.4795780>. <https://doi.org/10.1121/1.4795780>
14. B. Rafaely, *Fundamentals of Spherical Array Processing*, vol. 8. (Springer, Cham, Switzerland, 2015)
15. F. Zotter, M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. (Springer, Cham, Switzerland, 2019)
16. H. Møller, Fundamentals of binaural technology. *Appl Acoust.* **36**(3–4), 171–218 (1992)
17. Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, B. Rafaely, Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *J Acoust Soc Am.* **141**(6), 4087–96 (2017). <https://doi.org/10.1121/1.4983652>. <https://doi.org/10.1121/1.4983652>



18. C. Schörkhuber, M. Zauschirm, R. Höldrich, in *Proceedings of the DAGA*, vol. 44. Binaural rendering of ambisonic signals via magnitude least squares (German Acoustical Society (DEGA), Berlin, 2018), pp. 339–42
19. C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, I.J. Tashev, in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Improving binaural ambisonics decoding by spherical harmonics domain tapering and coloration compensation (IEEE, 2019), pp. 261–65
20. T. Lübeck, H. Helmholz, J.M. Arend, C. Pörschmann, J. Ahrens, Perceptual evaluation of mitigation approaches of impairments due to spatial undersampling in binaural rendering of spherical microphone array data. *J Audio Eng Soc.* **68**(6), 428–40 (2020)
21. T. Lübeck, H. Helmholz, J.M. Arend, C. Pörschmann, J. Ahrens, in *Proceedings of the 23-rd Int. Conf. on Digital Audio Effects (DAFx2020)*, vol. 1, ed. by G. Evangelista. Perceptual evaluation of mitigation approaches of impairments due to spatial undersampling in binaural rendering of spherical microphone array data: dry acoustic environments (J. Audio Eng. Soc, 2020), pp. 250–57
22. E. Rasumow, *Synthetic reproduction of head-related transfer functions by using microphone arrays. PhD thesis.* (Universität Oldenburg, Oldenburg, Germany, 2015)
23. E. Rasumow, M. Blau, S. Doclo, M. Hansen, D. Püschel, V. Mellert, et al, Perceptual evaluation of individualized binaural reproduction using a virtual artificial head. *J Audio Eng Soc.* **65**(6), 448–59 (2017)
24. M. Fallahi, M. Hansen, S. Doclo, S. van de Par, D. Püschel, M. Blau, Evaluation of head-tracked binaural auralizations of speech signals generated with a virtual artificial head in anechoic and classroom environments. *Acta Acustica.* **5**, 30 (2021)
25. L.S. Davis, R. Duraiswami, E. Grassi, N.A. Gumerov, Z. Li, D.N. Zotkin, in *Audio Engineering Society Convention 119*. High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues, (2005). <http://www.aes.org/e-lib/browse.cfm?elib=13369>
26. A.M. O'Donovan, D.N. Zotkin, R. Duraiswami. Spherical microphone array based immersive audio scene rendering (International Community for Auditory Display, 2008)
27. A. McKeag, D.S. McGrath, in *Audio Engineering Society Convention 6r*. Sound field format to binaural decoder with head tracking (Audio Engineering Society, 1996)
28. M. Noisternig, T. Musil, A. Sontacchi, R. Holdrich, in *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003*. 3D binaural sound reproduction using a virtual ambisonic approach, (2003), pp. 174–78. <https://doi.org/10.1109/VECIMS.2003.1227050>
29. P.G. Craven, M.A. Gerzon, *Coincident microphone simulation covering three dimensional space and yielding various directional outputs.* Google Patents. US Patent 4,042,779, (US patent, 1977)
30. W. Song, W. Ellermeier, J. Hald, Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise. *J Acoust Soc Am.* **123**(2), 910–24 (2008)
31. W. Song, W. Ellermeier, J. Hald, Psychoacoustic evaluation of multichannel reproduced sounds using binaural synthesis and spherical beamforming. *J Acoust Soc Am.* **130**(4), 2063–75 (2011)
32. P. Calamia, S. Davis, C. Smalt, C. Weston, in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017 IEEE Workshop On*. A conformal, helmet-mounted microphone array for auditory situational awareness and hearing protection (IEEE, 2017), pp. 96–100
33. J. Jiang, B. Xie, H. Mai, in *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*. The number of virtual loudspeakers and the error for spherical microphone array recording and binaural rendering (Audio Engineering Society, 2018)
34. L. Madmoni, J. Donley, V. Tourbabin, B. Rafaely, in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*. Beamforming-based binaural reproduction by matching of binaural signals (Audio Engineering Society, 2020)
35. H.L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory.* (Wiley, 2004)
36. D.L. Alon, B. Rafaely, *Spatial decomposition by spherical array processing. Parametric time-frequency domain spatial audio.* (Wiley, 2017), pp. 25–48
37. E.G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* (Academic press, 1999)
38. J. Meyer, G. Elko, in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2*. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield, (2002), pp. 1781–84. <https://doi.org/10.1109/ICASSP.2002.5744968>
39. J. Bitzer, K.U. Simmer, in *Microphone Arrays*. Superdirective microphone arrays (Springer, 2001), pp. 19–38
40. Z. Ben-Hur, J. Sheaffer, B. Rafaely, Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction. *Appl Acoust.* **134**, 138–44 (2018)
41. A. McLaren, Optimal numerical integration on a sphere. *Math Comput.* **17**(84), 361–83 (1963)
42. D.Y. Levin, E.A. Habets, S. Gannot, A generalized theorem on the average array directivity factor. *IEEE Sig Process Lett.* **20**(9), 877–80 (2013)
43. E. Gilbert, S. Morgan, Optimum design of directive antenna arrays subject to random variations. *Bell Syst Tech J.* **34**(3), 637–63 (1955)
44. Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, B. Rafaely, Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *J Acoust Soc Am.* **141**(6), 4087–96 (2017)
45. B. Rafaely, Phase-mode versus delay-and-sum spherical microphone array processing. *IEEE Sig Process Lett.* **12**(10), 713–16 (2005)
46. R.H. Hardin, N.J. Sloane, McLaren's improved snub cube and other new spherical designs in three dimensions. *Discrete Comput Geom.* **15**(4), 429–41 (1996)
47. B. Bernschütz, in *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics*. A spherical far field HRIR/HRTF compilation of the Neumann KU 100 (German Acoustical Society (DEGA), Berlin, 2013), p. 29
48. I.-R. Recommendation, *BS. 1534-3 : Method for the Subjective Assessment of Intermediate Quality Levels of Coding Systems.* (International Telecommunications Union, Geneva, 2015)
49. J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics. *J Acoust Soc Am.* **65**(4), 943–50 (1979)
50. P. Kabal, TSP speech database. McGill Univ Database Version. **1**(0), 09–02 (2002)
51. A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, S. Weinzierl, A spatial audio quality inventory (SAQI). *Acta Acustica U Acustica.* **100**(5), 984–94 (2014)
52. J.M. Chambers, *Graphical Methods for Data Analysis.* (CRC Press, New York, 2018)
53. T. Raykov, G.A. Marcoulides, *An Introduction to Applied Multivariate Analysis.* (Routledge, New York, 2008)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---