


## Article

# DOA Estimation of Indoor Sound Sources Based on Spherical Harmonic Domain Beam-Space MUSIC

Liuqing Weng<sup>1,2</sup>, Xiyu Song<sup>1,2,\*</sup> , Zhenghong Liu<sup>1,2</sup>, Xiaojuan Liu<sup>3</sup>, Haocheng Zhou<sup>1,2</sup>, Hongbing Qiu<sup>1,2</sup> and Mei Wang<sup>3</sup>

<sup>1</sup> Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, Guilin 541004, China

<sup>2</sup> School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China

<sup>3</sup> School of Information Science & Engineering, Guilin University of Technology, Guilin 541006, China

\* Correspondence: songxiyu@guet.edu.cn

**Abstract:** The Multiple Signal Classification (MUSIC) algorithm has become one of the most popular algorithms for estimating the direction-of-arrival (DOA) of multiple sources due to its simplicity and ease of implementation. Spherical microphone arrays can capture more sound field information than planar arrays. The collected multichannel speech signals can be transformed from the space domain to the spherical harmonic domain (SHD) for processing through spherical modal decomposition. The spherical harmonic domain MUSIC (SHD-MUSIC) algorithm reduces the dimensionality of the covariance matrix and achieves better DOA estimation performance than the conventional MUSIC algorithm. However, the SHD-MUSIC algorithm is prone to failure in low signal-to-noise ratio (SNR), high reverberation time (RT), and other multi-source environments. To address these challenges, we propose a novel joint spherical harmonic domain beam-space MUSIC (SHD-BMUSIC) algorithm in this paper. The advantage of decoupling the signal frequency and angle information in the SHD is exploited to improve the anti-reverberation property of the DOA estimation. In the SHD, the broadband beamforming matrix converts the SHD sound pressure to the beam domain output. Beamforming enhances the incoming signal in the desired direction and reduces the SNR threshold as well as the dimension of the signal covariance matrix. In addition, the 3D beam of the spherical array has rotational symmetry and its beam steering is decoupled from the beam shape. Therefore, the broadband beamforming constructed in this paper allows for the arbitrary adjustment of beam steering without the need to redesign the beam shape. Both simulation experiments and practical tests are conducted to verify that the proposed SHD-BMUSIC algorithm has a more robust adjacent source discrimination capability than the SHD-MUSIC algorithm.

**Keywords:** direction of arrival estimation; beam direction diagram; MUSIC; spherical harmonic domain; spatial spectrum



**Citation:** Weng, L.; Song, X.; Liu, Z.; Liu, X.; Zhou, H.; Qiu, H.; Wang, M. DOA Estimation of Indoor Sound Sources Based on Spherical Harmonic Domain Beam-Space MUSIC. *Symmetry* **2023**, *15*, 187. <https://doi.org/10.3390/sym15010187>

Academic Editor: Zine El Abidine Fellah

Received: 24 November 2022

Revised: 22 December 2022

Accepted: 3 January 2023

Published: 9 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The ability to localize acoustic events is a fundamental prerequisite for equipping microphones with awareness of their surrounding environment. Source localization provides estimates of positional information, e.g., Directions-of-Arrival (DOAs) [1]. Conventional DOA estimation methods are mainly based on time difference of arrival (TDOA) [2], steered response power—phase transform (SRP-PHAT) [3], beam scanning [4], and spatial covariance matrix (SCM) [5]—and so on. The high-resolution spectral estimation method, such as Multiple Signal Classification (MUSIC), has received much attention because it can break the Rayleigh limit constraint and has a more promising direction resolution. However, while handling coherent sources of broadband signals in a reverberant environment, the high-resolution spectral estimation method suffers from a large spectral peak search operation, sensitivity to array error, and a high-resolution SNR threshold. These shortcomings often have a great negative impact on research, which is worth considering deeply.

In recent years, the theory of spherical microphone arrays has been widely used and developed. For DOA estimation of multiple sound sources, the spherical microphone array [6–8] has become a key and universal research focus. Compared with a linear and planar array, the spherical array has rotational symmetry and is easy for beamforming [9,10]. The spherical microphone array enables sampling of non-confounding sound pressure to reconstruct the 3D sound field accurately. Furthermore, the sound pressure received can be transformed into the spherical harmonic domain (SHD) for processing [11]. Using the spherical Fourier transform (SFT), the spherical sound pressure received can be decomposed into several orders of spherical harmonic function combinations [12]. Thus, a MUSIC and Root-MUSIC pseudo-spectrum can be constructed in the SHD [13]. This spherical wave decomposition can separate the frequency and angular components of the sound source from each other. In [14], the authors used the frequency smoothing method to de-correlate the coherent sources, which improved the estimation performance of the SHD-MUSIC algorithm in the presence of coherent sources. However, it is clear that the signal subspace turns out not absolutely orthogonal to the noise subspace in the sound field environments. As a result, the SHD-MUSIC algorithm is sensitive to the effect of noise. There is a phenomenon called the subspace swap. The data covariance matrix cannot always hold at full-rank, which affects the correctness of the matrix decomposition results. The above reasons eventually reduce the performance of the DOA estimation algorithm in case of low SNRs. In the open spherical arrays [15], the zero point of the spherical Bessel function causes a low value of the spherical mode amplitude response [16]. At the Bessel zero points, the low amplitude problem of spherical modal amplitude cannot be responded to perfectly, so the direction of the signal source cannot be estimated. Two ideas can solve the Bessel zero problem: spherical configuration and algorithmic optimization. For example, we can use directional microphone arrays, rigid-sphere arrays [17], and dual-sphere arrays [18] under these circumstances. In [19], the authors proposed the SHD-RMUSIC algorithm using relative sound pressure values that are insensitive to noise to improve the robustness to noise. However, the algorithm could not achieve satisfactory performance at lower SNRs (below 5 dB) and was not applicable to strong reverberation (above 0.6 s) environments. To improve the performance of source direction estimation in strongly reverberant environments, the authors proposed the direct path dominance test (DPD-MUSIC) [20]. The DPD algorithm exploits the non-smoothness and sparsity of the speech signal. By selecting time-frequency bins (TF-bins) that contain the direct sound dominance of a target source without the remaining reflected sound contribution, these TF-bins are used to construct the spatial spectrum. Thus, a certain reverberation multi-path distortion problem is solved in some scenarios, and multiple sources can be localized in a 3D strongly reverberant environment. However, its application presupposes that most of the TF-bins candidate sets must contain correct DOA information and that the proportion of TF-bins containing correct DOA information is sufficient to estimate the sound source DOA reliably. To make this application prerequisite guaranteed, the authors proposed an improved DPD test method, called the DPD-EDS test [21]. The test identifies the TF-bins dominated by the direct sound by means of enhanced decomposition based on the direct sound. This improves the accuracy of selecting the TF-bins that contain the correct DOA information. Obviously, the DPD test requires a direct acoustic prior [22], which is often difficult to obtain in harsh sound field environments. This shows that the rigid spherical microphone arrays for indoor DOA estimation of multiple adjacent sources to attenuate the effects of low-SNR environments and Bessel zeros on DOA estimation performance are an effective avenue of research.

In this paper, a rigid spherical microphone array is used to perform DOA estimation of indoor sound sources by placing 32 omnidirectional microphones in a uniform sampling distribution on the surface of a rigid baffle. We propose a joint SHD-beamforming and beam-domain MUSIC algorithm solution. It is promising and profound that the proposed method can improve the resolution probability of adjacent sound sources and reduce the SNR resolution thresholds. In a nutshell, our contributions are the following:

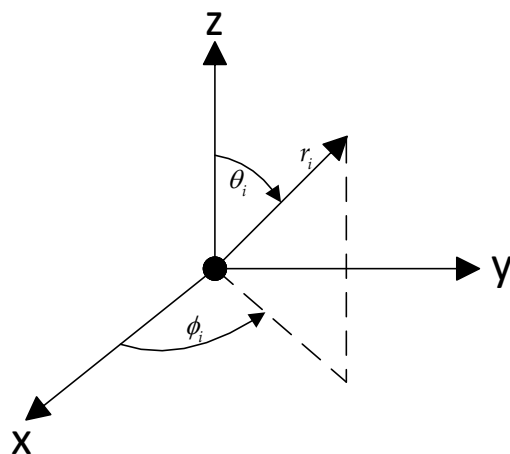
- We propose a novel spherical harmonic domain beam-space MUSIC algorithm. Combined with the advantages of the rigid-sphere configuration, the effects of the low SNR environment and Bessel's zero on the performance of multi-source DOA estimation are better reduced. This results in a more robust multi-source localization capability and adjacent source discrimination;
- We use a quadratic-constraint quadratic-planning optimization method to generate a multi-objective optimized signal-independent beam weight. We construct a very flexible rotationally symmetric beamformer so that the 3D beam of the spherical array can be arbitrarily redirected without redesigning the beam shape;
- We verify the superior performance of the proposed algorithm using simulated data as well as field-testing data in real-world situations (anechoic room and reverberant room).

The paper is structured as follows: Section 2 describes the space domain and spherical harmonic domain system models. Section 3 introduces the proposed DOA estimation method. Our experiment findings and discussions are clearly shown in Section 4. Finally, we complete the summary part and conclusions in Section 5.

## 2. System Models

### 2.1. Space-Domain System Model

We consider a set of spherical microphone arrays consisting of  $I$  omnidirectional microphones;  $r_i = (r_i \cos \phi_i \sin \theta_i, r_i \sin \phi_i \sin \theta_i, r_i \cos \theta_i)^T$  denotes the position of the  $i$ -th microphone of the array. The azimuth  $\phi$  is measured counterclockwise from the x-axis, the elevation angle  $\theta$  is measured downward from the z-axis, and  $r_i$  represents the distance of the  $i$ -th microphone to the center of the array. The adopted spherical coordinate system is shown in Figure 1.



**Figure 1.** Defined spherical coordinate system.

There are  $L$  far-field sound sources generating plane waves that propagate through space and are picked up by microphones.  $\Psi_l = (\theta_l, \phi_l)$  denotes the direction of propagation of the  $l$ -th sound source,  $k_l = -(k \cos \phi_l \sin \theta_l, k \sin \phi_l \sin \theta_l, k \cos \theta_l)^T$  denotes the wave number vector of the  $l$ -th plane wave. The signal received by the  $i$ -th microphone in the frequency domain can be expressed as:

$$p_i(k) = \sum_{l=1}^L v_i(k, \Psi_l) s_l(k) + n_i(k) \quad (1)$$

where  $v_i(k, \Psi_l)$  denotes the direction of propagation of the  $i$ -th microphone associated with the  $l$ -th plane wave;  $s_l(k)$  is the amplitude vector of the  $l$ -th plane wave;  $n_i(k)$  is the noise

received by the  $i$ -th microphone. The frequency domain received sound pressure model is expressed in matrix form as follows:

$$\mathbf{p}(k) = \mathbf{V}(k, \Psi)\mathbf{s}(k) + \mathbf{n}(k) \quad (2)$$

where  $\mathbf{V}(k, \Psi)$  is the  $I \times L$  dimensional direction matrix;  $\mathbf{s}(k) = [s_1(k), \dots, s_L(k)]^T$  is the  $L$  dimensional source signal vector;  $\mathbf{n}(k) = [n_1(k), \dots, n_I(k)]^T$  is the  $I$  dimensional zero-mean Gaussian white noise vector; and  $\mathbf{n}(k)$  is assumed to be uncorrelated with  $\mathbf{s}(k)$ . For an open-sphere array (microphones in a free field),  $\mathbf{v}(k, \Psi_I)$  is denoted as:

$$\mathbf{v}(k, \Psi_I) = [e^{-jk_1^T r_1}, e^{-jk_1^T r_2}, \dots, e^{-jk_1^T r_I}]^T \quad (3)$$

## 2.2. Spherical Harmonic Domain System Model

For the case of rigid spherical array with acoustic scattering [23], the time delay method cannot be used to obtain the sound pressure directly. The sound field modal synthesis method will be introduced to calculate the sound pressure response of the spherical array [24]. According to the Fourier acoustic principle, for a spherical array of order  $N$ , the sampled sound pressure of the microphone at position  $(r, \Omega)$ ,  $\Omega = (\theta, \phi)$  on the sphere and its spherical harmonic domain (SHD) are expressed as:

$$\begin{aligned} \mathbf{p}_{nm}(k, r) &= \sum_{i=1}^I a_i \mathbf{p}(k, r, \Omega_i) [Y_n^m(\Omega_i)]^* \\ \mathbf{p}(k, r, \Omega) &= \sum_{n=0}^N \sum_{m=-n}^n \mathbf{p}_{nm}(k, r) Y_n^m(\Omega) \end{aligned} \quad (4)$$

The spherical Fourier basis function  $Y_n^m(\Omega)$  is called the  $n$ -th order spherical harmonic function (SHF) of the degree of freedom,  $m$ :

$$Y_n^m(\Omega) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \theta) e^{im\phi} \quad (5)$$

where  $P_n^m$  denotes the concatenated Lejeune function;  $P_n^m(\cos \theta)$  reflects the effect of  $\theta$  on the operation state of the SHF; and  $e^{im\phi}$  reflects the effect of  $\phi$  on the operation state of the SHF. When sampling spherical sound pressure, a spherical sample  $(\theta_i, \phi_i)$  and a sampling weight  $a_i$  are given.

To extract the sound field spherical harmonics, the microphone positions are required to satisfy the weighted orthogonality condition. The three types of aliasing-free sampling for spherical array signal processing [25] are equal-angle sampling, Gaussian sampling, and near-uniform sampling. In order to reduce the amount of multi-channel data operation, we use the uniform sampling method which can achieve the minimum number of sampling points. It has a sampling point count of  $(N+1)^2$  and the sampling weight  $a_i = \frac{4\pi}{I}$  is set as a constant.

In the SHD, the frequency and the angular components of the source direction matrix can be separated from each other as follows [20]:

$$\mathbf{v}(k, \Psi_I) = \mathbf{y}^T(\Omega_i) \mathbf{B}(kr) \mathbf{y}^*(\Psi_I) \quad (6)$$

where  $\mathbf{y}(\Psi_I) = [Y_0^0(\Psi_I) Y_1^{-1}(\Psi_I) Y_1^0(\Psi_I) Y_1^1(\Psi_I) \dots Y_N^N(\Psi_I)]^T$  is a  $(N+1)^2$  dimensional vector of SHFs in the sound-source direction, consisting of SHFs of different orders;  $\mathbf{y}(\Omega_i)$  is a  $(N+1)^2$  dimensional vector of SHFs in the microphone direction, having the same form as  $\mathbf{y}(\Psi_I)$ ;  $\mathbf{B}(kr)$  is the modal intensity matrix, which turns out to be a  $(N+1)^2 \times (N+1)^2$  dimensional symmetric matrix composed of the spherical modal amplitude response  $b_n(kr)$ .

$b_n(kr)$  is constructed based on the structure of the array as for the open and rigid spherical arrays, respectively:

$$\begin{aligned} b_{n,open}(kr) &= 4\pi i^n j_n(kr) \\ b_{n,rigid}(kr) &= 4\pi i^n \left( j_n(kr) - \frac{j_n'(kr)}{h_n'(kr)} h_n(kr) \right) \end{aligned} \quad (7)$$

where  $j_n$  and  $h_n$  are the first class  $n$  order spherical Bessel function and the second class  $n$  order spherical Hankel function, respectively. For the rigid spherical array, the first term of  $b_n(kr)$  describes the incident sound field, which is the same as the open spherical array, and the second term of  $b_n(kr)$  describes the scattered sound field.

For a plane wave with amplitude  $s_l(k)$  and wave number vector  $(k, \Psi)$ ,  $\Psi = (\theta_l, \phi_l)$ , the sound pressure received at the  $i$ -th microphone can be expressed as:

$$p_l(k, r, \Omega_i) = y^T(\Omega_i) B(kr) y^*(\Psi_l) s_l(k) \quad (8)$$

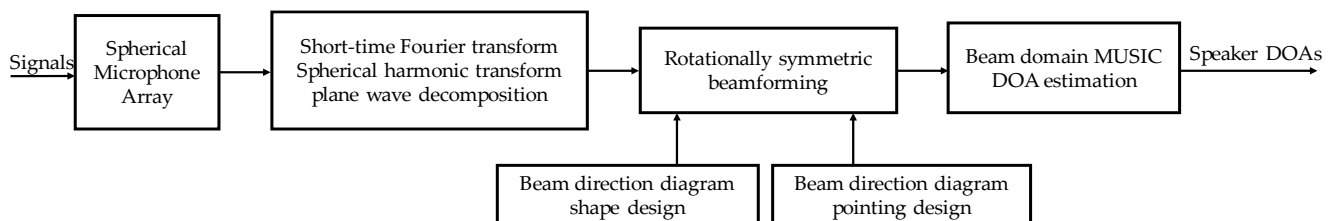
### 3. Methods

#### 3.1. Framework of the Proposed SHD-BMUSIC

It is known from the literature [19] that the SHD-MUSIC algorithm is sensitive to noise. To address these challenges, we introduce beam design in the SHD and transform the SHD-MUSIC algorithm into the beam space for processing; therefore, we combine the spherical harmonic domain (SHD) with the beam domain (BD). The beamforming is processed in the SHD and the MUSIC spatial spectrum is constructed in the BD. The main advantages are as follows:

1. The SHD axisymmetric beamformer is designed to separate the beam steering from the beam weights so that the beam steering can be adjusted arbitrarily without adjusting the beam weights;
2. The SHD beam weights can be independent of the signal frequency, without the demands on a special constant beam-width design;
3. The frequency and angle components of the source can be decoupled in the SHD, and the frequency smoothing can be used to decouple the coherent source without affecting the DOAs of the sources [14].

The framework of the proposed spherical harmonic domain beam-space MUSIC algorithm (SHD-BMUSIC) DOA estimation scheme is shown in Figure 2.



**Figure 2.** Flow chart of DOAs estimation scheme for indoor sound sources.

#### 3.2. Beam Weight Design

The beam weight design can be categorized as a single-objective and multi-objective optimization problem. The beam directional map metrics include directionality, white noise gain, side flap level, and main flap width. Although multi-objective formulation descriptions usually do not have closed-form solutions, they can be integrated into numerically solved optimization problems. The multi-objective beam weight design formulation is expressed in the form of a quadratic constraint quadratic programming (QCQP) optimization problem, which is a special case of second-order cone programming (SOCP):

$$\begin{aligned}
& \text{minimize} && \mathbf{w}_{nm}^H \mathbf{B} \mathbf{w}_{nm} \\
& \text{subject to} && \mathbf{w}_{nm}^H \mathbf{v}_{nm} = 1 \\
& && \mathbf{w}_{nm}^H \mathbf{A} \mathbf{w}_{nm} \leq \frac{1}{WNG_{\min}} \\
& \text{which} && \mathbf{A} = \mathbf{S} \mathbf{S}^H, \mathbf{S} = \frac{4\pi}{Q} \mathbf{Y}^H \\
& && \mathbf{B} = \frac{1}{4\pi} \text{diag}(|b_0|^2, |b_1|^2, |b_1|^2, |b_1|^2, \dots, |b_N|^2)
\end{aligned} \tag{9}$$

where  $WNG_{\min}$  is the lower bound on WNG. Matrix  $\mathbf{S}$  depends on the sampling scheme, for the near-uniform sampling scheme,  $\mathbf{S} = \frac{4\pi}{Q} \mathbf{Y}^H$ . The matrices  $\mathbf{A}$  and  $\mathbf{B}$  are conjugate symmetric matrices. For non-zero vectors  $\mathbf{w}_{nm}$ ,  $\mathbf{w}_{nm}^H \mathbf{A} \mathbf{w}_{nm}$ , and  $\mathbf{w}_{nm}^H \mathbf{B} \mathbf{w}_{nm}$  is positive definite.  $\mathbf{w}_{nm}^H \mathbf{A} \mathbf{w}_{nm} \leq \frac{1}{WNG_{\min}}$  denotes the white noise gain constraint.  $\mathbf{w}_{nm}^H \mathbf{v}_{nm} = 1$  denotes the distortion-free response constraint of the beam direction map in the array view direction. Judging by the satisfaction of distortion-free response constraint and white noise gain constraint, the average response  $\mathbf{w}_{nm}^H \mathbf{B} \mathbf{w}_{nm}$  in all directions of the array is minimized. The beamforming weight  $\mathbf{w}_{nm}$  is finally derived using a numerical solution method.

### 3.3. Beam Direction Chart Indicator

The directionality factor ( $DF$ ) is defined as the ratio of the observed directional signal power gain to the all-around uniform arrival signal power gain.

$$DF = \frac{|\mathbf{y}(\theta_k, \phi_k)|^2}{\frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi |\mathbf{y}(\theta, \phi)|^2 \sin \theta d\theta d\phi} \tag{10}$$

where  $\mathbf{y}(\theta_k, \phi_k)$  is the response output in the observation direction.  $DF$  quantifies the improvement in SNR provided by the array directional response. White noise gain (WNG) is defined as the improvement in SNR at the array output compared to the array input and is used as a measure of array robustness:

$$WNG = \frac{|\mathbf{w}_{nm}^H \mathbf{v}_{nm}|^2}{\mathbf{w}_{nm}^H \mathbf{S} \mathbf{S}^H \mathbf{w}_{nm}} \tag{11}$$

where  $\mathbf{w}_{nm}$  is the beamforming weight. Uniform sampling satisfies  $\mathbf{S} \mathbf{S}^H = \frac{4\pi}{Q} \mathbf{I}$ , where  $\mathbf{I}$  is the unit matrix and  $Q$  is the number of sampling samples.  $\mathbf{v}_{nm} = b_n(kr) [\mathbf{Y}_n^m(\theta_k, \phi_k)]^*$  denotes the array input generated by the plane wave sound field, and can be applied to other arrays such as rigid- or open-sphere arrays by simply modifying  $b_n(kr)$ . Therefore, the SHD-beamforming system has remarkable flexibility [26].

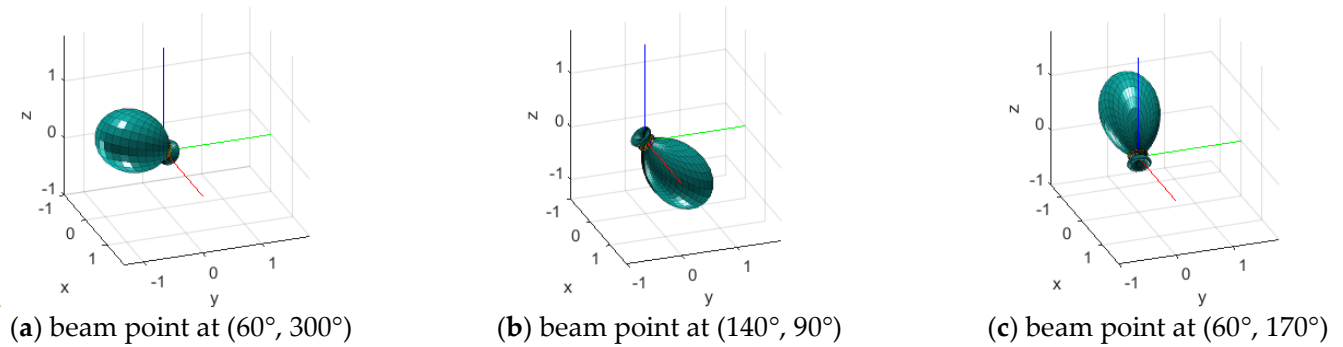
According to Equation (9), set the white noise gain constraint to  $\frac{1}{WNG_{\min}} = 0.2$  and 2.7 kHz frequency point. The beam weight value  $\mathbf{w}_{nm}$  is derived based on the QCQP algorithm, since  $\mathbf{w}_{nm}$  is independent of the signal frequency. Referring to  $\mathbf{w}_{nm}$ ,  $WNG = 22.95$  dB and  $DF = 24.0$  dB should be set for the entire operating frequency band.

### 3.4. Rotational Symmetric Beamformer

The SHD-beamforming consists of two parts: spherical harmonic decomposition and SHD weighting summation. If the beam steering angle is to be adjusted, the weighting vector needs to be redesigned. In the beamforming block diagram proposed by Meyer and Elko [27], reducing the beam formation weights to a one-dimensional function. The resulting beam direction map is axisymmetric when viewed from the viewpoint of the axis forming the symmetry. This separates the beam steering adjustment from the weighted value design:

$$\mathbf{w}_{nm, \text{sym}}^* = \frac{\mathbf{w}_{nm}}{b_n(kr)} \mathbf{Y}_n^m(\theta_0, \phi_0) \tag{12}$$

The rotationally symmetric beamformer contains two parts: beam guidance and beam synthesis. The weighted value  $w_{nm}$  adjusts the beam map shape and the spherical harmonics  $Y_n^m(\theta_0, \phi_0)$  adjust the beam steering angle. The rotationally symmetric beamformer with beam weight  $w_{nm}$  of Section 3.3 and different beam points  $Y_n^m(\theta_0, \phi_0)$  is shown in Figure 3.



**Figure 3.** Axisymmetric beam diagram with the same beam shape and different beam steering.

A plane wave decomposition (PWD) of  $p_{nm}(k, r)$ , i.e., dividing by the modal intensity  $B(kr)$ , yields the plane wave amplitude density can be expressed as:

$$\mathbf{a}_{nm}(k) = \mathbf{Y}^H(\Psi)\mathbf{s}(k) + \bar{\mathbf{n}}_{nm}(k) \quad (13)$$

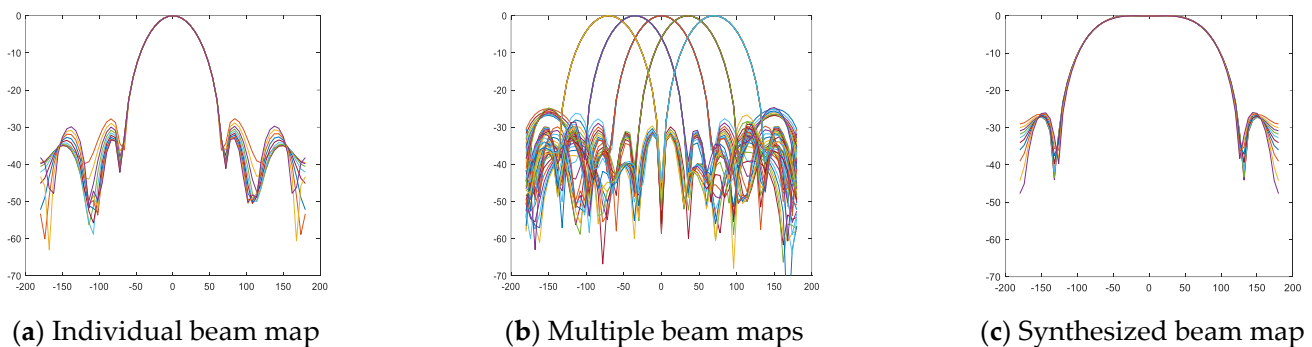
The beam response output is obtained using the inner product of axisymmetric beamforming weights  $w_{nm, sym}^*$  with SHD array data  $p_{nm}(k, r)$ :

$$\mathbf{y}(kr, \Delta\Psi) = \sum_{n=0}^N \sum_{m=-n}^n p_{nm}(k, r) w_{nm, sym}^* = \sum_{n=0}^N \sum_{m=-n}^n w_{nm} \mathbf{a}_{nm}(k) Y_n^m(\theta_0, \phi_0) \quad (14)$$

The beam response  $\mathbf{y}(kr, \Delta\Psi)$  is affected by the angle  $\Delta\Psi$  between the plane wave arrival direction  $(\theta_k, \phi_k)$  and the array view direction  $(\theta_0, \phi_0)$ .

### 3.5. Beam Domain MUSIC

The beam domain MUSIC algorithm forms multiple beams in the spatial sector of interest with the number of beams between the number of sources and the number of microphones [28]. Different pointing beams are used for synthesis to cover the target area, which forms a  $D \times (N + 1)^2$  dimensional axisymmetric beamforming matrix  $\mathbf{W}_{nm, sym} = [w_{nm1, sym}, \dots, w_{nmD, sym}]$ . The 2D preformed multi-beam matrix is shown in Figure 4.



**Figure 4.** Two-dimensional form of the preformed multi-beam directional map.

If the beamforming matrix  $W_{nm,sym}$  does not satisfy  $W_{nm,sym}^H W_{nm,sym} = I$ , the standard orthogonalization needs to be performed on it as follows:

$$\bar{W}_{nm,sym} = W_{nm,sym} (W_{nm,sym}^H W_{nm,sym})^{-1/2} \quad (15)$$

The beam response vector, obtained using the inner product of the beamforming weighting matrix and SHD array data, is expressed in matrix form:

$$\mathbf{y}(kr, \Delta\Psi) = \bar{W}_{nm,sym}^H \mathbf{p}_{nm}(k, r) \quad (16)$$

The beam domain data covariance matrix,  $\mathbf{S}_y(k) = E\{\mathbf{y}(kr, \Delta\Psi)\mathbf{y}^H(kr, \Delta\Psi)\}$ , is reduced to  $D \times D$  dimensions and the operations of the feature decomposition are reduced to  $o(D^3)$ . Therefore, the dimension-reduction beam-space processing eases the operations of the subspace algorithm.

Frequency smoothing (FS) can be applied as a straightforward average by assuming frequency independence of the (mode strength compensated) array manifold [29]. We de-correlate the coherent source signal by implementing FS that computes the smoothed covariance matrix as the average of covariance matrices at different frequency sectors. Then, the smoothed beam response covariance matrix  $\tilde{\mathbf{S}}_y = \frac{1}{K} \sum_{q=1}^K \mathbf{S}_y(k_q)$  is obtained.

The eigenvalue decomposition of  $\tilde{\mathbf{S}}_y$  is performed to obtain the noise subspace and the signal subspace in the beam domain. The BD-MUSIC spatial pseudo-spectral function is constructed using the orthogonality of the beam scan vector  $\mathbf{b}(\Psi) = \bar{W}_{nm,sym}^H \mathbf{y}(\Psi)$  and beam domain noise subspace  $E_{nm}$ . The equations can be summarized as follows:

$$P_{SHD-BMUSIC}(\Psi) = \frac{\mathbf{b}^H(\Psi)\mathbf{b}(\Psi)}{\mathbf{b}^H(\Psi)\bar{W}_{nm,sym}^H E_{nm} (\bar{W}_{nm,sym}^H E_{nm})^H \mathbf{b}(\Psi)} \quad (17)$$

Finally, the spectral peak search is carried out and the direction corresponding to the peak is the DOA of the sound sources.

## 4. Experiment

### 4.1. Simulation Parameter Settings

We simulated the room impulse response (RIRs) between an omnidirectional source and a microphone in a reverberant environment using the image-source method modified to account for the scattering of a rigid sphere [23]. Multiple randomly selected audios from the TIMIT database were used as a simulated source of the original vocalizations. The original audio signal was a pure voice signal of 4 s duration with a sampling rate of 16 kHz. The indoor 3D empty room size was set to 4 m  $\times$  6 m  $\times$  3 m. The radius of the sphere array  $r$  was 0.042 m, placed in the middle of the room, and the distance of the original sound sources from the center of the sphere array was 1 m. The spherical array was configured as a rigid sphere using uniform sampling weights. The order  $N$  of the spherical harmonic function was taken as three.

The signals received of the sphere array were obtained by convolving the original audio with the RIRs and by adding a certain SNR of Gaussian white noise. The signals received were transformed to the time–frequency domain using the short-time Fourier transform (STFT). The frame length of STFT was 16 ms, and the selected window function was a Hamming window with a window length of 256 points and a frame overlap of 50%. Then, we carried out a spherical harmonic transform in the used frequency band. A disadvantage of rigid spherical arrays is the low-frequency performance, so excessive noise amplification at lower frequencies due to modal strength compensation needs to be avoided [29]. According to the spatial aliasing  $kr < N$ , the upper-frequency limit is around 5 kHz. Thirty frequency bins ranging from 2700 Hz to 3600 Hz were



used to perform the spherical harmonic transform to obtain the SHD sound pressure data. Finally, DOA estimation algorithms were executed separately to obtain the DOAs of the sound sources. The azimuthal accuracy and elevation angle accuracy of the spectrum peak search were 1 degree. The localization performances of the SHD-MUSIC and SHD-BMUSIC are compared in these different SNR levels and reverberation time (RT) environments.

#### 4.2. DOA Estimated Evaluation Metrics

To evaluate the algorithm fairly, the performances of the algorithm using two qualitative metrics were measured. Fifty Monte Carlo simulation trials were performed in each sound field condition.

The first metric was the average number of detected sources, which analyzed whether the DOA estimation algorithm was supposed to detect the target source. For each detection, the target source was detected if the deviation between the estimated angle and the true angle was set within 20°. The number of sources detected each time was statistically averaged to obtain the average number of detected sources.

When multiple sources can be correctly distinguished, we used a second evaluation metric of the average root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{m=1}^M \sum_{l=1}^L (|\theta_{ori}^m(l) - \theta_{est}^m(l)| + |\phi_{ori}^m(l) - \phi_{est}^m(l)|)^2}{LM}} \quad (18)$$

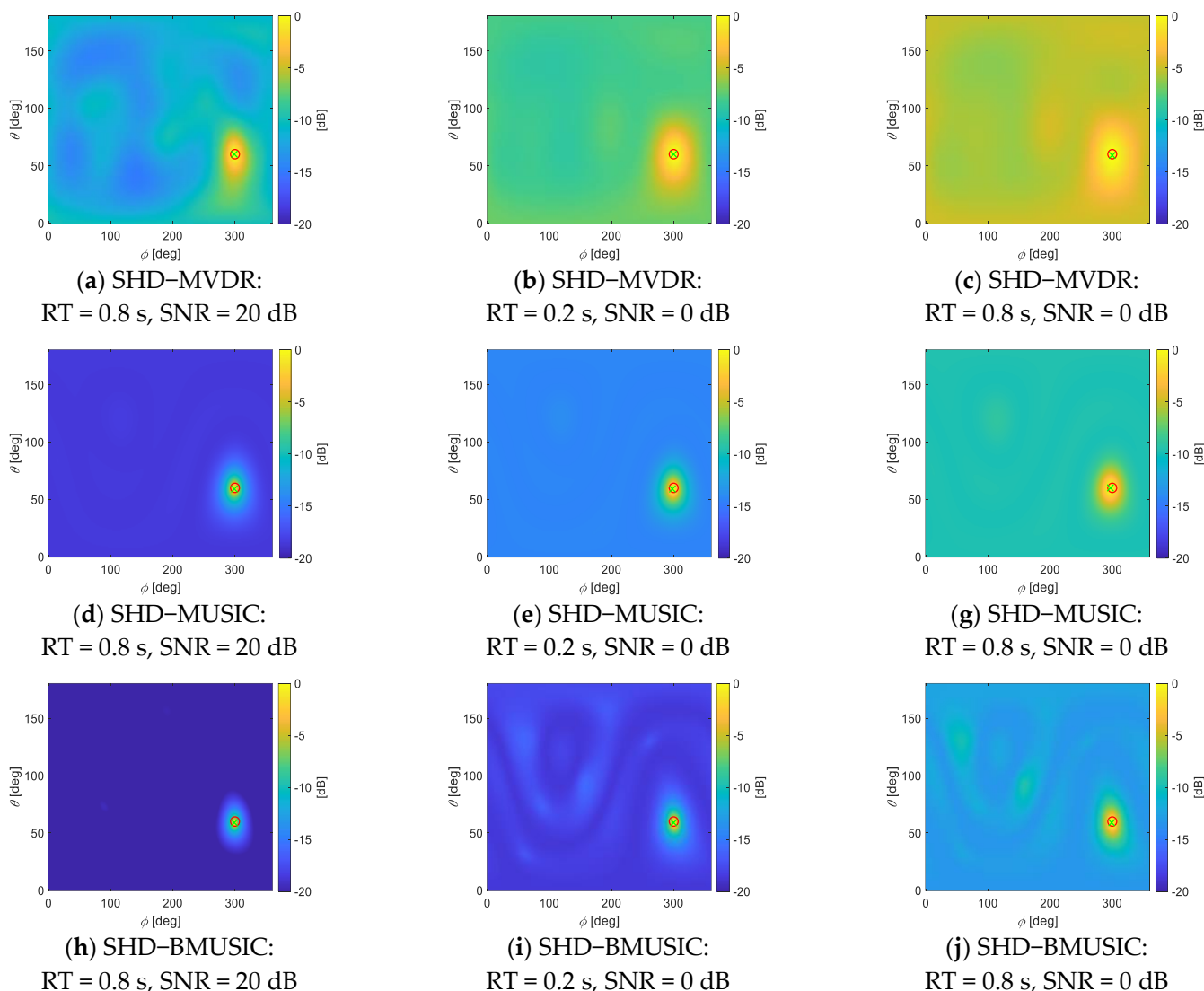
where  $L$  is the number of sound sources;  $M$  is the number of cases that successfully detect all the  $L$  sound sources;  $(\theta_{ori}^m(l), \phi_{ori}^m(l))$  is the true position of the  $l$ -th source in the  $m$ -th trial;  $(\theta_{est}^m(l), \phi_{est}^m(l))$  is the estimated position of the  $l$ -th source in the  $m$ -th trial. The average of the estimated positions of all trials is taken as the final estimated position  $(\bar{\theta}_{est}(l), \bar{\phi}_{est}(l))$  of the  $l$ th source. The average RMSE determines the deviation between the true position and the estimated position and is used to analyze the accuracy of the localization algorithm. Only the RMSE of the positioned angle of the detected signal is considered.

#### 4.3. Simulation Testing Results

##### 4.3.1. Single-Source Localization Algorithm Verification

First, the estimation results of the algorithms for spherical harmonic domain-minimum variance distortion-free response (SHD-MVDR) [12], SHD-MUSIC, and SHD-BMUSIC are compared for the single-source case. Aiming at minimizing the contribution of noise and any arriving signals from other directions, the MVDR method is supposed to maintain a fixed gain in the look direction. Three groups of single-source experiments with different sound field environments were tested. The sound source was located at (60°, 300°). The sound field conditions were high reverberation (RT = 0.8 s, SNR = 20 dB), low SNR (RT = 0.2 s, SNR = 0 dB), and high-reverberation low-SNR (RT = 0.8 s, SNR = 0 dB). For the SHD-BMUSIC algorithm, five beams were formed to cover the area where the target sound source was located. The estimated spatial spectrum of the single source for the three algorithms is shown in Figure 5.

The spatial spectrum of the single source showed that all three algorithms successfully estimated the location of the single source in different environments. However, the spatial spectral peaks of SHD-BMUSIC were both sharper than SHD-MUSIC and SHD-MVDR and the spatial spectral gain was higher. It was quite convincing that the SHD-BMUSIC algorithm suppressed the noise outside the beam region under these circumstances and the SNR outputs were improved.

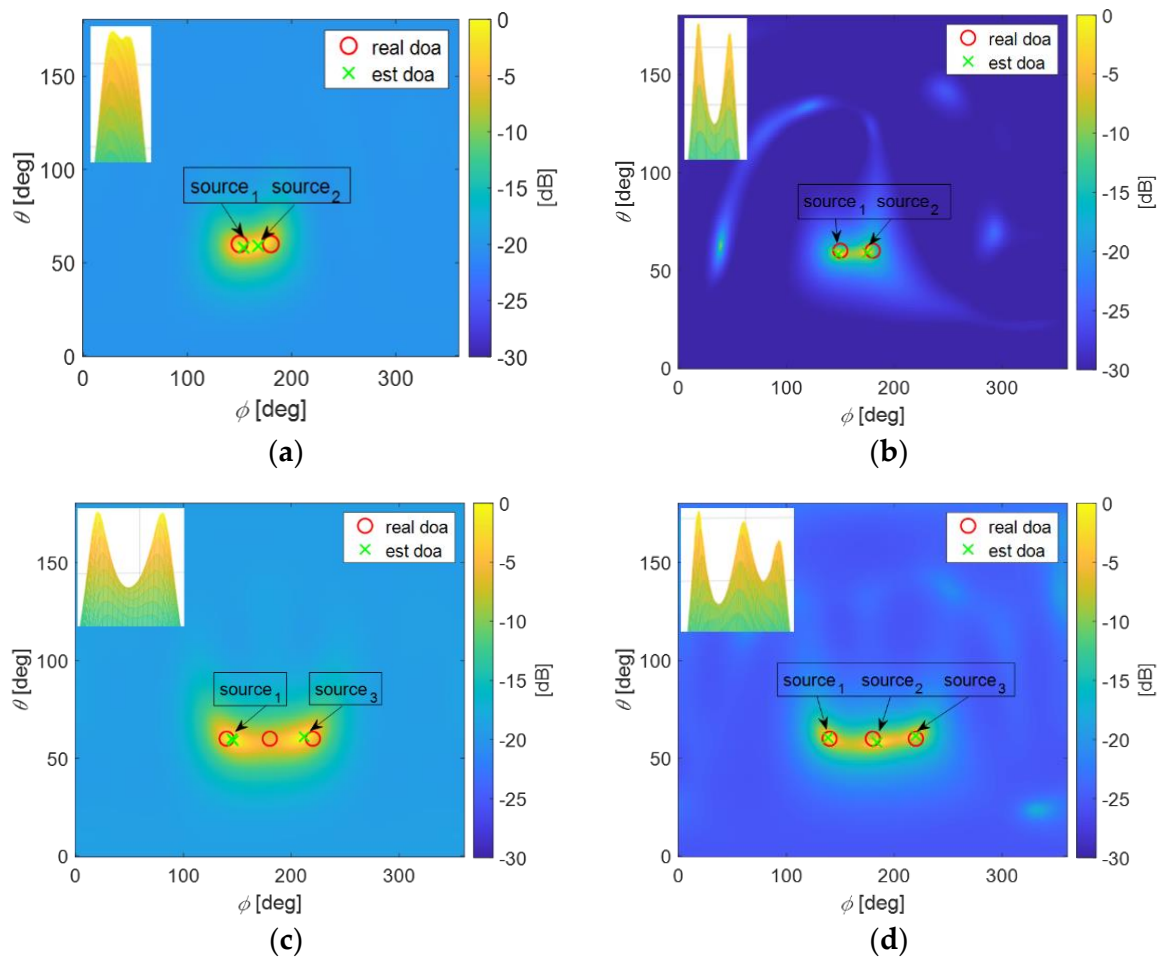


**Figure 5.** Spatial spectra of single sound source in different acoustic environments.

#### 4.3.2. Adjacent Sound Sources Localization Algorithm Verification

In this section, we verify that the SHD-BMUSIC was able to distinguish the locations of multiple adjacent sound sources with higher source identification accuracy compared to SHD-MUSIC. Two sets of experiments with the adjacent sound sources were performed. The first set of experiments set up sound source 1 at  $(60^\circ, 150^\circ)$  and sound source 2 at  $(60^\circ, 180^\circ)$ . The second group of experiments set up sound source 1 at  $(60^\circ, 140^\circ)$ , sound source 2 at  $(60^\circ, 180^\circ)$ , and sound source 3 at  $(60^\circ, 220^\circ)$ . Both experiments were conducted in an environment with  $RT = 0.3$  s and  $SNR = 20$  dB. For the SHD-BMUSIC algorithm, for the first set of experiments, five beams were used to cover the area where two sources were located. For the second set of experiments, seven beams were used to cover the area where the three sources were located. The spatial spectra of the adjacent sound sources for both algorithms are shown in Figure 6.

It is quite convincing that the algorithm proposed has a better discriminative ability than SHD-MUSIC. Tables 1 and 2 show the experimental statistics for the two cases, respectively.



**Figure 6.** Spatial spectra of adjacent sound sources at SNR = 20 dB, RT = 0.3 s. (a) SHD–MUSIC: two adjacent sound sources; (b) SHD–BMUSIC: two adjacent sound sources; (c) SHD–MUSIC: three adjacent sound sources; (d) SHD–BMUSIC: three adjacent sound sources.

**Table 1.** Localization results and errors for two adjacent sound sources at SNR = 20 dB, RT = 0.3 s.

Methods	$(\bar{\theta}_{est}(1), \bar{\phi}_{est}(1))$	$(\bar{\theta}_{est}(2), \bar{\phi}_{est}(2))$	Average Number of Detected Sources	Average RMSE
SHD-MUSIC	(58°, 155°)	(59°, 171°)	1.45	5.4261
SHD-BMUSIC	(59°, 149°)	(58°, 173°)	1.96	4.6440

**Table 2.** Localization results and errors for three adjacent sound sources at SNR = 20 dB, RT = 0.3 s.

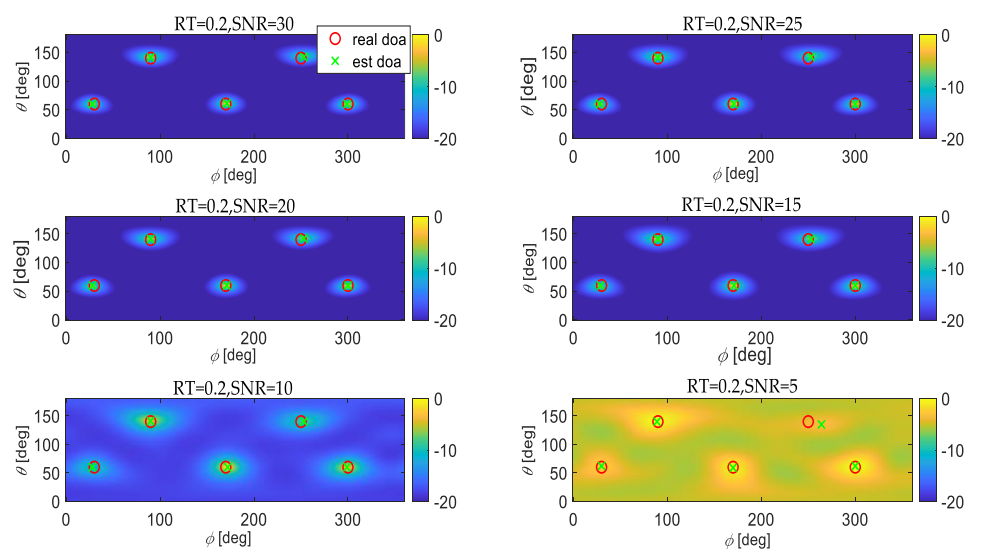
Methods	$(\bar{\theta}_{est}(1), \bar{\phi}_{est}(1))$	$(\bar{\theta}_{est}(2), \bar{\phi}_{est}(2))$	$(\bar{\theta}_{est}(3), \bar{\phi}_{est}(3))$	Average Number of Detected Sources	Average RMSE
SHD-MUSIC	(60°, 145°)	(59°, 147°)	(61°, 212°)	2.06	3.4549
SHD-BMUSIC	(60°, 139°)	(58°, 185°)	(62°, 221°)	2.84	2.8284

As shown in Tables 1 and 2, for the first set of conditions, both algorithms succeeded most of the time in distinguishing two adjacent sound sources. However, SHD-BMUSIC had a higher average number of detected sources and a lower average RMSE. For the second set of conditions, SHD-MUSIC was basically unable to completely estimate the three adjacent sound sources. The intermediate source was not successfully detected in

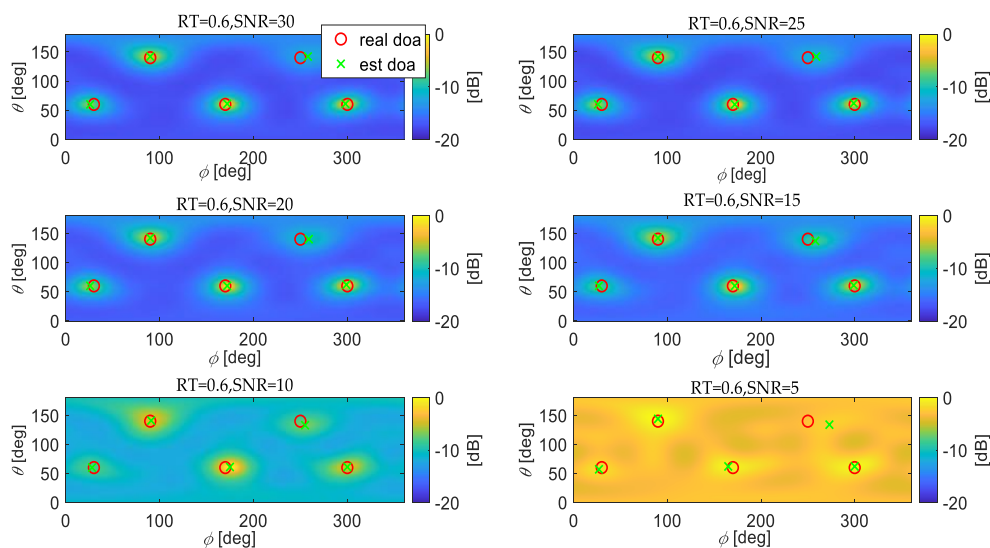
most cases, resulting in a small average number of detected sources for SHD-MUSIC. In contrast, SHD-BMUSIC was able to detect all three sound sources in most tests. In the tests where all sound sources were successfully discriminated, SHD-BMUSIC also had a smaller average RMSE. As a result, SHD-BMUSIC improved the accuracy of sound source location estimation and had a higher adjacent source discrimination capability. The increase in position resolution was the gain from the beamforming operation.

### 4.3.3. Multi-Source Localization Algorithm Verification at Different SNRs and RTs

To verify the performance of the SHD-BMUSIC algorithm in the reverberation-noise sound field, we considered the positioning results under different reverberation durations and SNR levels. There were five sources emitting simultaneously, and the source locations were  $(60^\circ, 300^\circ)$ ,  $(140^\circ, 90^\circ)$ ,  $(60^\circ, 170^\circ)$ ,  $(140^\circ, 250^\circ)$ , and  $(60^\circ, 30^\circ)$ . There were 25 beams used to form five new separate beams to the source area and to cover the target area. The spatial spectra of SHD-BMUSIC at different SNRs and RTs are shown in Figure 7.



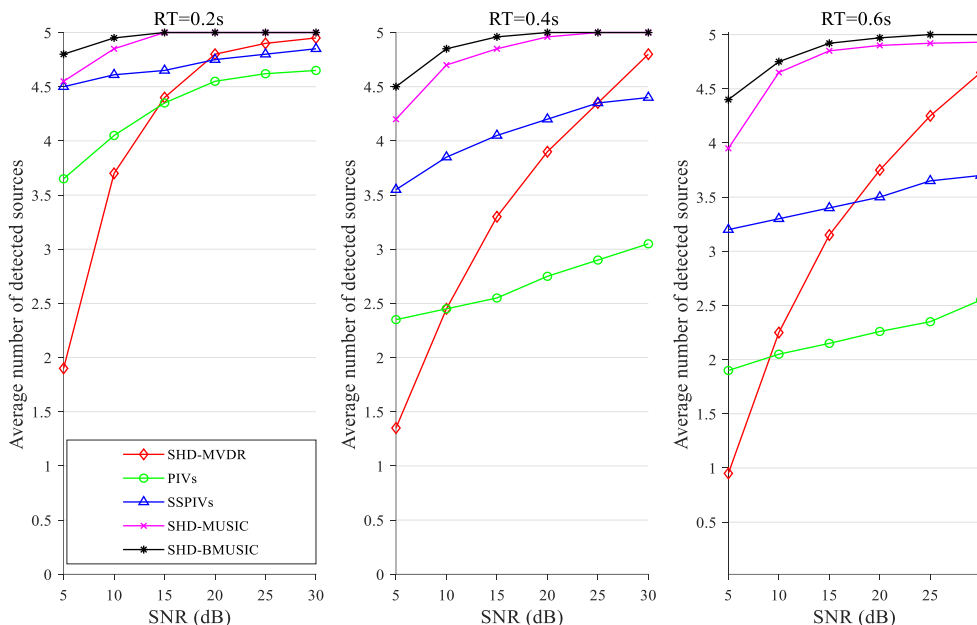
(a) Spatial spectra of SHD-BMUSIC at different SNRs and low RT



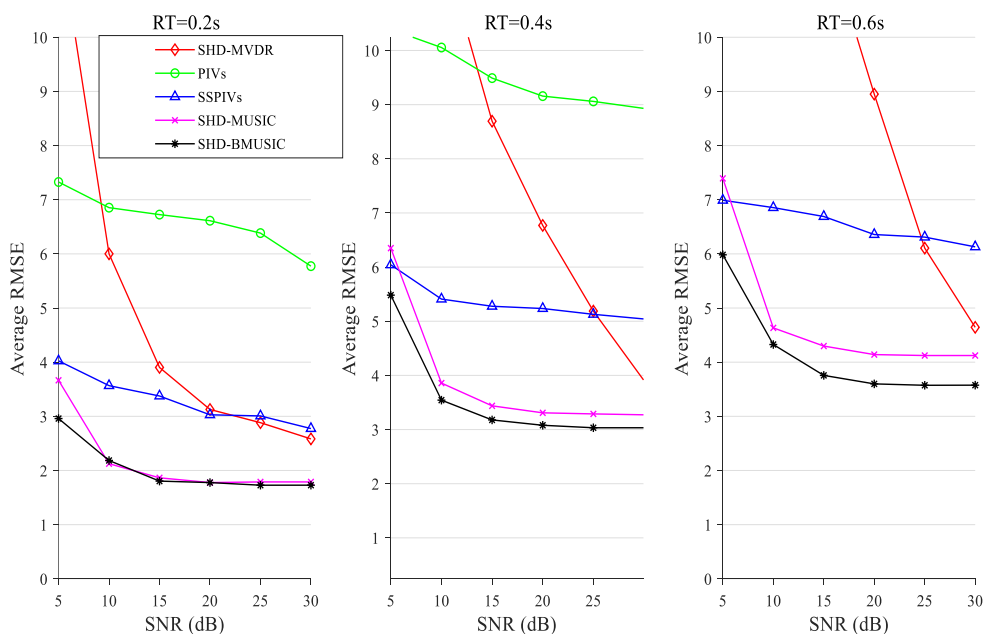
(b) Spatial spectra of SHD-BMUSIC at different SNRs and high RT

Figure 7. Multi-source spatial spectra of SHD-BMUSIC at different SNRs and RTs (five sources).

The statistical results of SHD-MVDR, PIVs [29], SSPIVs [29], SHD-MUSIC, and SHD-BMUSIC algorithms were compared for multi-source case testing. Figure 8 shows the average results of the tests with five sources for each combination of SNR and RT.



(a) Average number of detected sources with different SNRs and RTs of localization algorithms



(b) Average RMSE with different SNRs and RTs of localization algorithms

Figure 8. Line graph of the statistical trials of the localization algorithm (five sources).

As seen in Figure 8, overall, the performance of the localization algorithm showed a certain degree of decline as the SNR decreased and the RT increased. The performance of SHD-MVDR was seriously affected by reverberation and noise. SSPIVs improved the shortcomings of PIVs which were severely affected by reverberation. PIVs and SSPIVs had shorter running times because they were not expected to carry out a spectral peak search. The SHD-MUSIC and SHD-BMUSIC, which use frequency smoothing, had strong

adaptability to reverberation. Due to the gain from beamforming, the proposed algorithm had better sound-source detection capability and localization accuracy than the rest of the algorithms, showing better anti-noise and anti-reverberation performance.

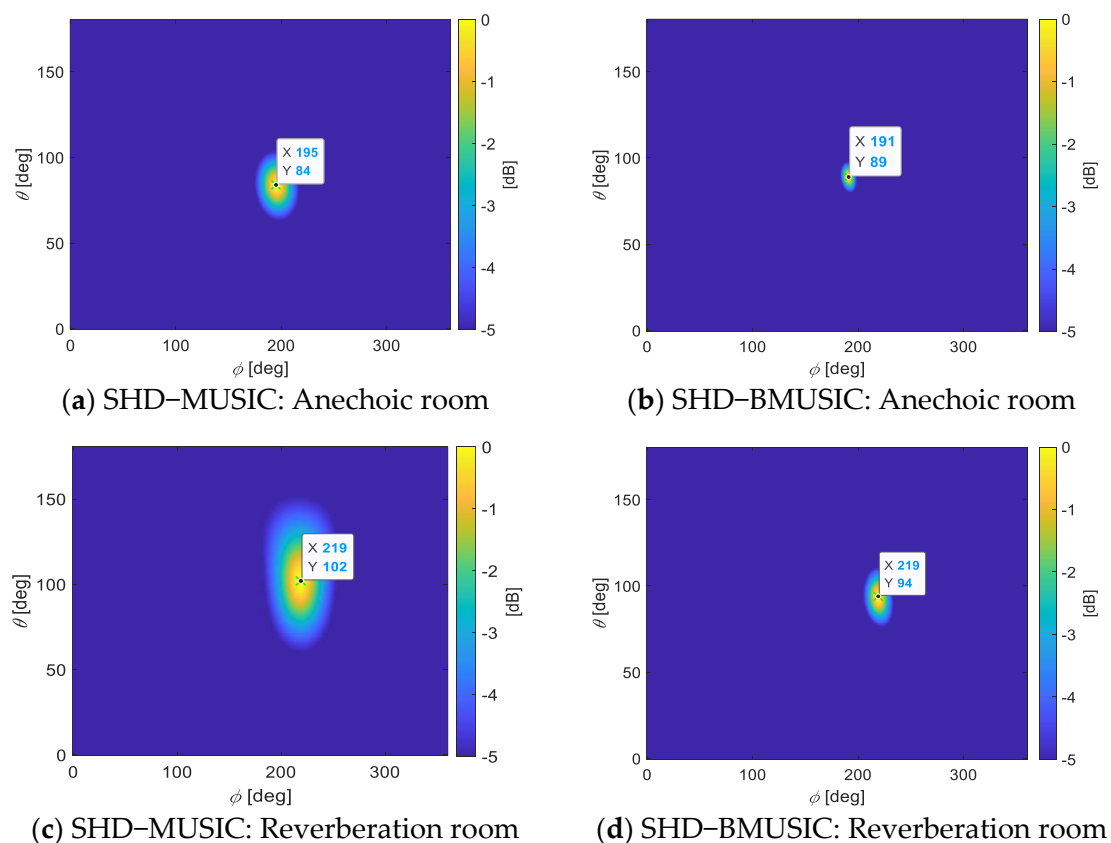
#### 4.4. Field Testing Results

The spherical microphone equipment was used to collect the recording data in a standard anechoic and reverberant room. Each audio segment was recorded for 4 s and the rest of the parameter settings were the same as the simulation parameter settings. The actual spherical microphone array we used and the actual experimental room for field testing are shown in Figure 9.



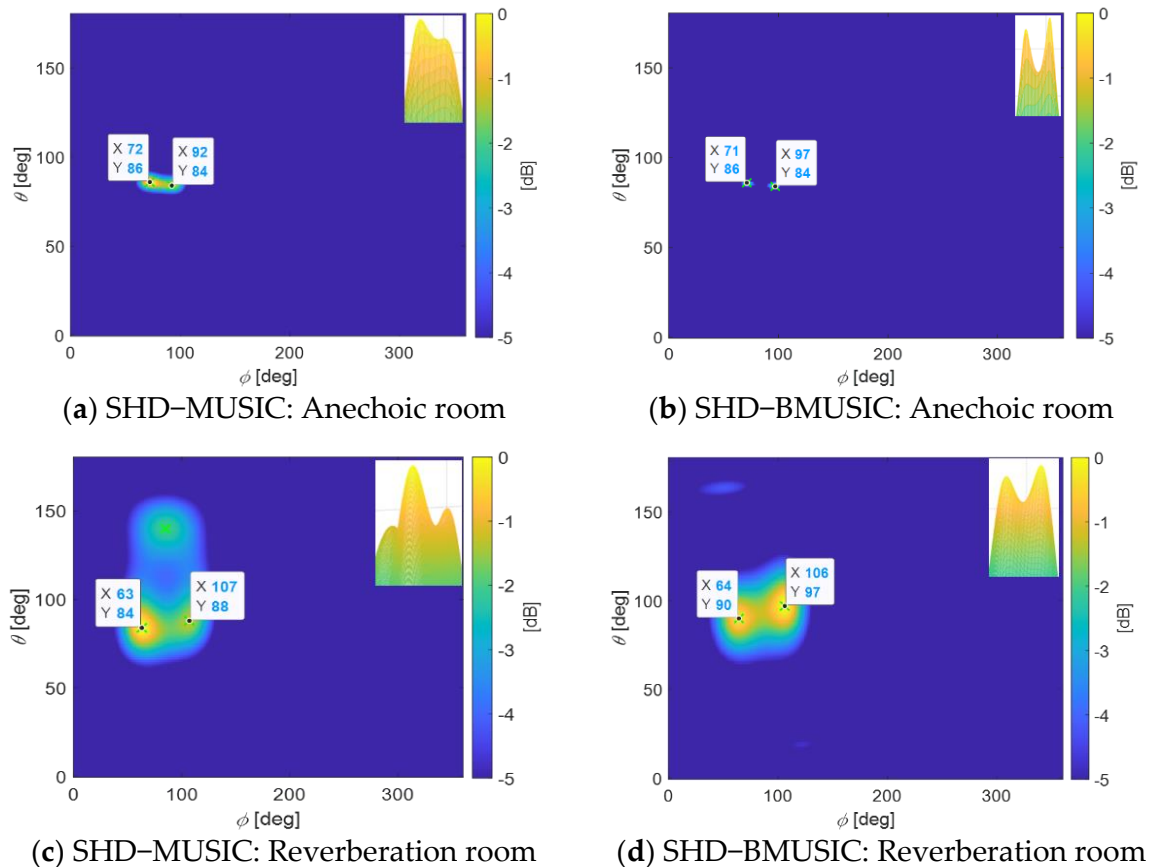
**Figure 9.** Experimental equipment and different field-testing scenarios.

First of all, the case of the single-sound source was tested. The source in the anechoic room was located at  $(90^\circ, 190^\circ)$ ; the source in the reverberant room was located at  $(90^\circ, 220^\circ)$ . The spatial spectra estimated by the two algorithms are shown in Figure 10.



**Figure 10.** Spatial spectra for actual testing of single-sound source.

Then, the case of the adjacent dual-sound sources was tested. The locations of the two sources in the anechoic and reverberant rooms were  $(90^\circ, 70^\circ)$  and  $(90^\circ, 100^\circ)$ . The estimated spatial spectra of the two algorithms are shown in Figure 11.



**Figure 11.** Spatial spectra for actual testing of adjacent dual-sound sources.

Finally, it is promising to conclude that both algorithms are better at estimating the location of a single source in an anechoic room. The SHD-BMUSIC had sharper peaks and lower estimation errors. In the reverberation room, SHD-MUSIC showed additional excess peaks due to reverberant reflected sound in adjacent dual-source experiments. This made it easy for the algorithm to mistake the pseudo-peaks for the direction of the sound source, leading to an incorrect estimation of the sound source location. However, SHD-BMUSIC did not have the phenomenon of redundant pseudo-peaks. This was because the beamforming suppressed the excess reflected sound outside the source region, thus improving the reverberation resistance of the algorithm.

## 5. Conclusions

In this paper, we use a spherical harmonic domain beam-space MUSIC algorithm to solve the multi-source DOA estimation problem in harsh environments where reverberation and noise coexist. The designed spherical harmonic-domain beam is not only easy to implement, with flexible adjustment of steering, but also reduces the matrix dimension to reduce the amount of eigen-decomposition operations. From the experimental results, it can be seen that the SHD-BMUSIC algorithm has some advantages over the SHD-MUSIC algorithm without the beam domain strategy. Its offline construction of the beam does not increase the time consumption of online real-time positioning. Instead, it enables online localization with sharper spatial spectral peaks, more robust multi-source localization capability, and adjacent source discrimination. We will further optimize the design in the direction of beamforming and the MUSIC algorithm to better solve the problem of

locating and tracking multiple adjacent sound sources in high-reverberation and high-noise environments.

**Author Contributions:** Conceptualization, L.W. and X.S.; software, L.W, X.L. and H.Z.; formal analysis M.W.; writing—review and editing, L.W, X.S. and H.Q.; funding acquisition, H.Q., Z.L. and X.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Natural Science Foundation of China: 62071135, the Guangxi Natural Science Foundation: 2019GXNSFBA245103, and Project (CRKL200111) from the Key Laboratory of Cognitive Radio and Information Processing, Ministry of Education (Guilin University of Electronic Technology).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. The data can be found on <https://en.wikipedia.org/wiki/TIMIT> (accessed on 1 March 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Evers, C.; Lollmann, H.W.; Mellmann, H.; Schmidt, A.; Barfuss, H.; Naylor, P.A.; Kellermann, W. The LOCATA Challenge: Acoustic Source Localization and Tracking. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 1620–1643. [[CrossRef](#)]
2. He, H.; Wu, L.; Lu, J.; Qiu, X.; Chen, J. Time Difference of Arrival Estimation Exploiting Multichannel Spatio-Temporal Prediction. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 463–475. [[CrossRef](#)]
3. Diaz-Guerra, D.; Miguel, A.; Beltran, J.R. Robust Sound Source Tracking Using SRP-PHAT and 3D Convolutional Neural Networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 300–311. [[CrossRef](#)]
4. Jaafer, Z.; Goli, S.; Elameer, A.S. Performance Analysis of Beam Scan, MIN-NORM, Music and Mvdr DOA Estimation Algorithms. In Proceedings of the 2018 International Conference on Engineering Technology and their Applications (IICETA), Al-Najaf, Iraq, 8–9 May 2018; pp. 72–76.
5. Xu, C.; Xiao, X.; Sun, S.; Rao, W.; Chng, E.S.; Li, H. Weighted Spatial Covariance Matrix Estimation for MUSIC Based TDOA Estimation of Speech Source. In Proceedings of the Interspeech 2017, ISCA, Stockholm, Sweden, 20–24 August 2017; pp. 1894–1898.
6. Tervo, S.; Politis, A. Direction of Arrival Estimation of Reflections from Room Impulse Responses Using a Spherical Microphone Array. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1539–1551. [[CrossRef](#)]
7. Hu, Y.; Lu, J.; Qiu, X. Direction of Arrival Estimation of Multiple Acoustic Sources Using a Maximum Likelihood Method in the Spherical Harmonic Domain. *Appl. Acoust.* **2018**, *135*, 85–90. [[CrossRef](#)]
8. Choi, J.-W.; Zotter, F.; Jo, B.; Yoo, J.-H. Multiarray Eigenbeam-ESPRIT for 3D Sound Source Localization With Multiple Spherical Microphone Arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 2310–2325. [[CrossRef](#)]
9. Yadav, S.K.; George, N.V. Sparse Distortionless Modal Beamforming for Spherical Microphone Arrays. *IEEE Signal Process. Lett.* **2022**, *29*, 2068–2072. [[CrossRef](#)]
10. Yan, S.; Sun, H.; Svensson, U.P.; Ma, X.; Hovem, J.M. Optimal Modal Beamforming for Spherical Microphone Arrays. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 361–371. [[CrossRef](#)]
11. Kumar, L.; Hegde, R.M. Near-Field Acoustic Source Localization and Beamforming in Spherical Harmonics Domain. *IEEE Trans. Signal Process.* **2016**, *64*, 3351–3361. [[CrossRef](#)]
12. Khaykin, D.; Rafaely, B. Acoustic Analysis by Spherical Microphone Array Processing of Room Impulse Responses. *J. Acoust. Soc. Am.* **2012**, *132*, 261–270. [[CrossRef](#)]
13. Kumar, L.; Bi, G.; Hegde, R.M. The Spherical Harmonics Root-Music. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 3046–3050.
14. Khaykin, D.; Rafaely, B. Coherent Signals Direction-of-Arrival Estimation Using a Spherical Microphone Array: Frequency Smoothing Approach. In Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 17–20 October 2009; pp. 221–224.
15. Pan, X.; Wang, H.; Lou, Z.; Su, Y. Fast Direction-of-Arrival Estimation Algorithm for Multiple Wideband Acoustic Sources Using Multiple Open Spherical Arrays. *Appl. Acoust.* **2018**, *136*, 41–47. [[CrossRef](#)]
16. Samarasinghe, P.; Abhayapala, T.; Poletti, M. Wavefield Analysis Over Large Areas Using Distributed Higher Order Microphones. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 647–658. [[CrossRef](#)]
17. Coteli, M.B.; Hacihabiboglu, H. Multiple Sound Source Localization with Rigid Spherical Microphone Arrays via Residual Energy Test. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 790–794.
18. Jin, C.T.; Epain, N.; Parthy, A. Design, Optimization and Evaluation of a Dual-Radius Spherical Microphone Array. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 193–204. [[CrossRef](#)]



19. Hu, Y.; Abhayapala, T.D.; Samarasinghe, P.N. Multiple Source Direction of Arrival Estimations Using Relative Sound Pressure Based MUSIC. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 253–264. [[CrossRef](#)]
20. Nadiri, O.; Rafaely, B. Localization of Multiple Speakers under High Reverberation Using a Spherical Microphone Array and the Direct-Path Dominance Test. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 1494–1505. [[CrossRef](#)]
21. Madmoni, L.; Rafaely, B. Direction of Arrival Estimation for Reverberant Speech Based on Enhanced Decomposition of the Direct Sound. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 131–142. [[CrossRef](#)]
22. Rafaely, B.; Alhaiany, K. Speaker Localization Using Direct Path Dominance Test Based on Sound Field Directivity. *Signal Processing* **2018**, *143*, 42–47. [[CrossRef](#)]
23. Jarrett, D.P.; Habets, E.A.P.; Thomas, M.R.P.; Naylor, P.A. Rigid Sphere Room Impulse Response Simulation: Algorithm and Applications. *J. Acoust. Soc. Am.* **2012**, *132*, 1462–1472. [[CrossRef](#)]
24. Hu, Y.; Samarasinghe, P.N.; Abhayapala, T.D.; Dickins, G. Modeling Characteristics of Real Loudspeakers Using Various Acoustic Models: Modal-Domain Approaches. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 561–565.
25. Rafaely, B.; Weiss, B.; Bachmat, E. Spatial Aliasing in Spherical Microphone Arrays. *IEEE Trans. Signal Process.* **2007**, *55*, 1003–1010. [[CrossRef](#)]
26. Wang, L.; Zhu, J. Flexible Beampattern Design Algorithm for Spherical Microphone Arrays. *IEEE Access* **2019**, *7*, 139488–139498. [[CrossRef](#)]
27. Meyer, J.; Elko, G. A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, Orlando, FL, USA, 13–17 May 2002; pp. II-1781–II-1784.
28. Li, X.; Yan, S.; Ma, X.; Hou, C. Spherical Harmonics MUSIC versus Conventional MUSIC. *Appl. Acoust.* **2011**, *72*, 646–652. [[CrossRef](#)]
29. Moore, A.H.; Evers, C.; Naylor, P.A. Direction of Arrival Estimation in the Spherical Harmonic Domain Using Subspace Pseudointensity Vectors. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 178–192. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.