

Article Binaural Heterophasic Superdirective Beamforming

Yuzhu Wang ¹, Jingdong Chen ^{1,*}, Jacob Benesty ², Jilu Jin ¹, and Gongping Huang ³

- ¹ Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an 710072, China; im.yuzhu.wang@gmail.com (Y.W.); charles.jilu.jin@gmail.com (J.J.)
- ² INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Montreal, QC H5A 1K6, Canada; benesty@emt.inrs.ca
- ³ Andrew and Erna Viterby Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel; gongping@campus.technion.ac.il
- * Correspondence: jingdongchen@ieee.org

Abstract: The superdirective beamformer, while attractive for processing broadband acoustic signals, often suffers from the problem of white noise amplification. So, its application requires well-designed acoustic arrays with sensors of extremely low self-noise level, which is difficult if not impossible to attain. In this paper, a new binaural superdirective beamformer is proposed, which is divided into two sub-beamformers. Based on studies and facts in psychoacoustics, these two filters are designed in such a way that they are orthogonal to each other to make the white noise components in the binaural beamforming outputs incoherent while maximizing the output interaural coherence of the diffuse noise, which is important for the brain to localize the sound source of interest. As a result, the signal of interest in the binaural superdirective beamformer's outputs is in phase but the white noise components in the outputs are random phase, so the human auditory system can better separate the acoustic signal of interest from white noise by listening to the outputs of the proposed approach. Experimental results show that the derived binaural superdirective beamformer is superior to its conventional monaural counterpart.

Keywords: microphone arrays; binaural beamforming; heterophasic; superdirective beamformer; white noise gain; directivity factor; beampattern; interaural coherence

1. Introduction

Microphone arrays combined with proper beamforming methods have been used in a wide range of applications, such as hearing aids, smart headphones, smart speakers, voice communication, automatic speech recognition (ASR), human–machine interfaces, etc., to extract signals of interest from noisy observations. Many beamformers have been developed over the last few decades [1–17], among which the so-called superdirective beamformer [9] is particularly attractive. It is derived by maximizing the directivity factor (DF), which is equivalent to maximizing the gain in signal-to-noise ratio (SNR) in diffuse noise, subject to the distortionless constraint at the endfire direction. So, this beamformer is more efficient than other fixed beamformers for suppressing noise, interference, and reflections incident from different directions in noisy and reverberant environments [14,18]. It also has a frequency-invariant beampattern if the sensor spacing is small, which is essential for acquiring high-fidelity broadband acoustic and speech signals.

However, there is one major drawback with the existing superdirective beamforming approach: the white noise amplification is very serious at low frequencies. As a consequence, the application of this beamformer requires well-designed arrays with high-quality microphones of extremely low self-noise level, e.g., at least below 0 dB(A) for second- and higher-order superdirective beamformers, which is difficult if not impossible to attain. This white noise amplification problem considerably limits the use of the superdirective beamformer in practical applications [18,19] and how to deal with this problem has become an important issue that has attracted a significant amount of research attention. A number of



Citation: Wang, Y.; Chen, J.; Benesty, J.; Jin, J.; Huang, G. Binaural Heterophasic Superdirective Beamforming. *Sensors* **2021**, *21*, 74. https://dx.doi.org/10. 3390/s21010074

Received: 31 October 2020 Accepted: 22 December 2020 Published: 25 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 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/).

2 of 22

methods have been developed subsequently in the literature, including the so-called robust superdirective beamformer [9,20], the combined superdirective beamformer [21], the optimized superdirective beamformer [22], the subspace superdirective beamformer [11,23], and the reduced-rank superdirective beamformer [24]. While the problem is approached from different perspectives, the fundamental principle underlying those methods stays the same, i.e., making a compromise between the DF and the level of white noise gain (WNG) [10,20,22,25,26]. In other words, all those methods attempt to circumvent the white noise amplification problem by sacrificing the DF and, as a consequence, the resulting beamformer may no longer be superdirective.

In this work, we take a different avenue. Instead of sacrificing the DF to improve the WNG, we design the superdirective beamformer, which consists of two sub-beamformers, each generates an output. The two sub-beamformers are designed to be orthogonal to each other so that the acoustic signal of interest in the binaural outputs is in phase while the (amplified) white noise is random phase. This design is strongly motivated by studies and facts in psychoacoustics, which showed that the location (or direction) information of signals has a significant impact on speech intelligibility in the human auditory system.

Many experiments have been conducted to study the influence of the direction information of speech and white noise (especially at frequencies below 1 kHz) on speech intelligibility [27–48]. Briefly, the impact of the source direction on the perception of speech in the human binaural auditory system can be classified into two scenarios: in phase and out of phase; while the perception of noise can be divided into three scenarios: in phase, random phase, and out of phase, where in phase means that in every frequency the binaural (two-channel) signals have the same phase and out of phase means that in every frequency the phase difference between the binaural signals is exactly 180°. An illustration of the impact of binaural signals and noise phase on speech intelligibility, inspired from [49], is shown in Figure 1, where the leftmost column indicates the phase relationship of the white noise at the left and right channels, and the first row represents the phase relationship of binaural speech signals. The combination indicates the influence of interaural direction relations on the localization of speech and white noise in space. The circle shape indicates that the signal is concentrated in a limited area, in front of the head for the case of in phase or on either side of the ears for the case of random phase. The rectangle shape indicates that the signal spreads in the area behind both ears in the case of out of phase.

The blue circle indicates that the speech signals at the left and right channels are in phase but the white noise components at the two channels are in random phase, which is related to the proposed binaural superdirective beamformer.

Listening tests have been conducted to study the intelligibility of different phase combinations. It is confirmed that the phase combination of the desired speech signal and white noise has a significant influence on intelligibility. A list of the most common six scenarios, summarized in [49] and represented in Table 1, can be divided into three categories: antiphasic, heterophasic, and homophasic. Listening tests showed that the antiphasic category corresponds to the highest intelligibility, which can be 25% higher than that of the homophasic category in low SNR scenarios. The intelligibility of the heterophasic scenario. Inspired by this, it is desirable to design beamformers that make the binaural outputs corresponding to the antiphasic or heterophasic cases.



Figure 1. Illustration of different phase scenarios and the influence of the interaural phase relations on the localization of speech and white noise in space. The circle shape means that the signal is concentrated in a limited area, in front of the head for the case of in phase or on either side of the ears for the case of random phase. The rectangle shape indicates that the signal spreads in the area behind the ears, which is related to the case of out of phase. The red dotted circle indicates that the speech signal and the white noise are both in phase, which is related to the monaural superdirective beamformer. The blue circle indicates that the speech signal is in phase while the white noise is in random phase, which is related to the binaural superdirective beamformer developed in this work. This figure is a modified version of the results in [49].

Table 1. Different scenarios for intelligibility study based on phase relationship between speech and noise [49].

Scenario	Speech	Noise	Class
1	Out of phase	In phase	Antiphasic
2	In phase	Out of phase	Antiphasic
3	In phase	Random phase	Heterophasic
4	Out of phase	Random phase	Heterophasic
5	In phase	In phase	Homophasic
6	Out of phase	Out of phase	Homophasic

In this paper, we use the interaural coherence (IC) to describe the auditory localization information. Consider a diffuse noise field; when the IC of the binaural signals reaches its maximum, i.e., 1, there is a precise region of the sound source, which is located in the middle of the head, i.e., the in-phase case; however, when the binaural signals are completely incoherent, i.e., the IC equals to 0, there are two independent sources at the two ears. This corresponds to the random phase case. In many approaches in the literature, binaural processing generates two collinear filters. The resulting output ICs for white (due to noise amplification) and diffuse noises are 1. So, both the signal of interest and the noise in the human auditory system are perceived to be in the same region. Consequently, our brain will have difficulties to separate them with a binaural presentation and intelligibility will certainly be affected. Apparently, the conventional monaural superdirective beamformer and the conventional collinear binaural processing belong to the homophasic case, which has the lowest intelligibility. To improve intelligibility, we propose a binaural superdirective beamformer by constructing two orthogonal filters, with which the IC for the white noise

components is equal to zero while the IC for the desired signal components is equal to one in the binaural outputs. Consequently, with the proposed method, our auditory system can more easily distinguish between the signal of interest and the (amplified) white noise by listening to the binaural outputs, leading to improved intelligibility [50]. An illustration of the proposed binaural superdirective beamformer with a uniform linear array is shown in Figure 2, which can suppress spatial noise while seperating the desired signal and the white noise into different perception zone. Since the binaural superdirective beamformer developed in this paper corresponds to the heterophasic case [49] (see also Table 1), we name it binaural heterophasic superdirective beamformer. Note that the IC information has been used in the traditional binaural speech enhancement methods for binaural cues preservation [51,52]. However, the proposed binaural heterophasic superdirective beamformer uses the IC magnitude information in a very different way where the two orthogonal sub-beamformers are designed to minimize the IC magnitude of the white noise components while maximizing the IC magnitude of the diffuse noise components in the binaural outputs to achieve better perceptual separation of the signal of interest and white noise.



Figure 2. Illustration of the proposed binaural superdirective beamformer, which suppresses directional acoustic interference and noise and meanwhile separates the desired signal and the white noise into different zones.

The rest of this paper is organized as follows. In Section 2, we present the signal model and formulate the problem. In Section 3, we briefly review the derivation of the conventional superdirective beamformer. In Section 4, we discuss binaural linear filtering and the associated performance measures. In Section 5, we derive the binaural heterophasic superdirective beamformer. Then, in Section 6, we present some experiments to validate the theoretical study. Finally, conclusions are presented in Section 7.

2. Signal Model and Problem Formulation

We consider a source signal of interest (plane wave), in the farfield, that propagates from the azimuth angle, θ , in an anechoic acoustic environment at the speed of sound, i.e., c = 340 m/s, and impinges on a uniform linear array (ULA) consisting of 2*M* omnidirectional microphones. In this scenario, the corresponding steering vector (of length 2*M*) is [5]

$$\mathbf{d}(\omega,\theta) = \begin{bmatrix} 1 & e^{-j\omega\tau_0\cos\theta} & \dots & e^{-j(2M-1)\omega\tau_0\cos\theta} \end{bmatrix}^T,$$
(1)

where *j* is the imaginary unit with $j^2 = -1$, $\omega = 2\pi f$ is the angular frequency, with f > 0 being the temporal frequency, $\tau_0 = \delta/c$ is the delay between two successive sensors at the angle $\theta = 0$, with δ being the interelement spacing, and the superscript ^{*T*} is the transpose operator.

Assume that the desired signal comes from a specific direction $\theta = \theta_s$. From the steering vector defined in (1), we can express the frequency-domain observation signal vector of length 2*M* as [2]

$$\mathbf{y}(\omega) = \begin{bmatrix} Y_1(\omega) & Y_2(\omega) & \cdots & Y_{2M}(\omega) \end{bmatrix}^T$$

= $\mathbf{x}(\omega) + \mathbf{v}(\omega)$
= $\mathbf{d}(\omega, \theta_s) X(\omega) + \mathbf{v}(\omega),$ (2)

where $Y_m(\omega)$ is the *m*th microphone signal, $\mathbf{x}(\omega) = \mathbf{d}(\omega, \theta_s)X(\omega)$, $X(\omega)$ is the zero-mean source signal of interest, which is also called the desired signal, $\mathbf{d}(\omega, \theta_s)$ is the signal propagation vector, which is same as the steering vector at $\theta = \theta_s$, and $\mathbf{v}(\omega)$ is the zero-mean additive noise signal vector defined similarly to $\mathbf{y}(\omega)$. We deduce that the $2M \times 2M$ covariance matrix of $\mathbf{y}(\omega)$ is

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{y}}(\omega) &\stackrel{\triangle}{=} & E\Big[\mathbf{y}(\omega)\mathbf{y}^{H}(\omega)\Big] \\ &= & \phi_{\mathbf{X}}(\omega)\mathbf{d}(\omega,\theta_{\mathbf{s}})\mathbf{d}^{H}(\omega,\theta_{\mathbf{s}}) + \mathbf{\Phi}_{\mathbf{v}}(\omega) \\ &= & \phi_{\mathbf{X}}(\omega)\mathbf{d}(\omega,\theta_{\mathbf{s}})\mathbf{d}^{H}(\omega,\theta_{\mathbf{s}}) + \phi_{V_{1}}(\omega)\Gamma_{\mathbf{v}}(\omega), \end{aligned}$$
(3)

where the superscript ^{*H*} is the conjugate-transpose operator, $E[\cdot]$ denotes mathematical expectation, $\phi_X(\omega) \stackrel{\triangle}{=} E[|X(\omega)|^2]$ is the variance of $X(\omega)$, $\Phi_{\mathbf{v}}(\omega) \stackrel{\triangle}{=} E[\mathbf{v}(\omega)\mathbf{v}^H(\omega)]$ is the covariance matrix of $\mathbf{v}(\omega)$, $\phi_{V_1}(\omega) \stackrel{\triangle}{=} E[|V_1(\omega)|^2]$ is the variance of the noise, $V_1(\omega)$, at the first sensor, and $\Gamma_{\mathbf{v}}(\omega) \stackrel{\triangle}{=} \Phi_{\mathbf{v}}(\omega)/\phi_{V_1}(\omega)$ is the pseudo-coherence matrix of the noise. We assume that noises at different sensors have the same variance.

In order to design the superdirective beamformer, we make two basic assumptions [9,19].

- (i) The sensor spacing, δ , is much smaller than the acoustic wavelength, $\lambda = c/f$, i.e., $\delta \ll \lambda$ (this implies that $\omega \tau_0 \ll 2\pi$). This assumption is required so that the true acoustic pressure differentials can be approximated by finite differences of the microphones' outputs.
- (ii) The desired source signal propagates from the angle $\theta_s = 0$ (endfire direction). Therefore, (2) becomes

$$\mathbf{y}(\omega) = \mathbf{d}(\omega, 0)X(\omega) + \mathbf{v}(\omega),\tag{4}$$

and, at the endfire, the value of the beamformer beampattern should always be equal to 1 (or maximal).

Our objective in this paper is to derive a binaural superdirective beamformer, which can take advantage of the human binaural auditory system to separate the desired speech signal from white noise so that the intelligibility of the beamformer's output signals will be higher than that of the output of the conventional (monaural) superdirective beamformer. To that end, we will find two various and useful estimates of $X(\omega)$, each for one of the binaural channels, so that along with our binaural hearing system, white noise amplification will be perceptually attenuated thanks to this binaural presentation.

3. Conventional Superdirective Beamformer

The conventional linear fixed beamforming technique is performed by applying a complex weight at the output of each microphone and then sum all the weighted outputs together to get an estimate of the source signal [2,19], i.e.,

...

$$Z(\omega) = \mathbf{h}^{H}(\omega)\mathbf{y}(\omega)$$

= $X(\omega)\mathbf{h}^{H}(\omega)\mathbf{d}(\omega,0) + \mathbf{h}^{H}(\omega)\mathbf{v}(\omega),$ (5)

where $Z(\omega)$ is the estimate of the desired source signal, $X(\omega)$, and $\mathbf{h}(\omega)$ is a spatial linear filter of length 2*M* containing all the complex weights. We see from (5) that the distortionless constraint should be

$$\mathbf{h}^{H}(\omega)\mathbf{d}(\omega,0) = 1.$$
(6)

Now, we can define the directivity factor (DF) of the beamformer as [3,19,43]

$$\mathcal{D}[\mathbf{h}(\omega)] \stackrel{\triangle}{=} \frac{|\mathbf{h}^{H}(\omega)\mathbf{d}(\omega,0)|^{2}}{\frac{1}{2}\int_{0}^{\pi}|\mathbf{h}^{H}(\omega)\mathbf{d}(\omega,\theta)|^{2}\sin\theta d\theta}$$
$$= \frac{|\mathbf{h}^{H}(\omega)\mathbf{d}(\omega,0)|^{2}}{\mathbf{h}^{H}(\omega)\Gamma_{\mathbf{d}}(\omega)\mathbf{h}(\omega)},$$
(7)

where $\Gamma_{d}(\omega) = \frac{1}{2} \int_{0}^{\pi} \mathbf{d}(\omega, \theta) \mathbf{d}^{H}(\omega, \theta) \sin \theta d\theta$ whose *ij*th (*i*, *j* = 1, 2, ..., 2M) element is

$$[\mathbf{\Gamma}_{d}(\omega)]_{ij} = \frac{\sin[\omega(j-i)\tau_{0}]}{\omega(j-i)\tau_{0}}$$

= sinc[\omega(j-i)\tau_{0}]. (8)

The matrix $\Gamma_d(\omega)$ can be viewed as the pseudo-coherence matrix of the spherically isotropic (diffuse) noise.

By taking into account the distortionless constraint in (6), the maximization of the DF in (7) leads to the well-known superdirective beamformer [9,53]:

$$\mathbf{h}_{\rm SD}(\omega) = \frac{\mathbf{\Gamma}_{\rm d}^{-1}(\omega)\mathbf{d}(\omega,0)}{\mathbf{d}^{H}(\omega,0)\mathbf{\Gamma}_{\rm d}^{-1}(\omega)\mathbf{d}(\omega,0)},\tag{9}$$

whose DF is

$$\mathcal{D}[\mathbf{h}_{\mathrm{SD}}(\omega)] = \mathbf{d}^{H}(\omega, 0) \mathbf{\Gamma}_{\mathrm{d}}^{-1}(\omega) \mathbf{d}(\omega, 0), \tag{10}$$

which, obviously, is maximal. Besides maximizing the DF, the other great advantage of $\mathbf{h}_{SD}(\omega)$ is that the corresponding beampattern is almost frequency invariant. However, white noise amplification is a tremendous problem. Consequently, the superdirective beamformer can only be used with a very small number of microphones and/or with regularization of the matrix $\Gamma_d(\omega)$, but this regularization affects the DF as well as the shape of the beampattern, which makes the beamformer more frequency dependent. Therefore, there is still a great interest to find new ideas to improve this superdirective beamformer.

4. Binaural Linear Filtering and Performance Measures

In this section, we explain binaural linear filtering in connection with fixed beamforming and propose some important performance measures in this context. The extension of the conventional (monaural) fixed linear beamforming to the binaural case can be done by applying two complex-valued linear filters, $\mathbf{h}_1(\omega)$ and $\mathbf{h}_2(\omega)$ of length 2*M*, to the observed signal vector, $\mathbf{y}(\omega)$, i.e.,

$$Z_{i}(\omega) = \mathbf{h}_{i}^{H}(\omega)\mathbf{y}(\omega)$$

$$= X(\omega)\mathbf{h}_{i}^{H}(\omega)\mathbf{d}(\omega,0) + \mathbf{h}_{i}^{H}(\omega)\mathbf{v}(\omega), i = 1, 2,$$
(11)

where $Z_1(\omega)$ and $Z_2(\omega)$ are two different estimates of $X(\omega)$. The variance of $Z_i(\omega)$ is then

$$\begin{aligned} \phi_{Z_{i}}(\omega) &= \mathbf{h}_{i}^{H}(\omega) \mathbf{\Phi}_{\mathbf{y}}(\omega) \mathbf{h}_{i}(\omega) \\ &= \phi_{X}(\omega) \left| \mathbf{h}_{i}^{H}(\omega) \mathbf{d}(\omega, 0) \right|^{2} + \mathbf{h}_{i}^{H}(\omega) \mathbf{\Phi}_{\mathbf{v}}(\omega) \mathbf{h}_{i}(\omega) \\ &= \phi_{X}(\omega) \left| \mathbf{h}_{i}^{H}(\omega) \mathbf{d}(\omega, 0) \right|^{2} + \phi_{V_{1}}(\omega) \mathbf{h}_{i}^{H}(\omega) \mathbf{\Gamma}_{\mathbf{v}}(\omega) \mathbf{h}_{i}(\omega). \end{aligned}$$
(12)

It is clear that the two distortionless constraints are

$$\mathbf{h}_{i}^{H}(\omega)\mathbf{d}(\omega,0) = 1, \ i = 1,2.$$
 (13)

A very important performance measure is the input SNR, which can be obtained from (3), i.e.,

$$iSNR(\omega) = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)}.$$
(14)

According to (12), the binaural output SNR can be defined as

$$\operatorname{oSNR}[\mathbf{h}_{1}(\omega),\mathbf{h}_{2}(\omega)] = \frac{\phi_{X}(\omega)}{\phi_{V_{1}}(\omega)} \times \frac{\sum_{i=1}^{2} \left| \mathbf{h}_{i}^{H}(\omega) \mathbf{d}(\omega,0) \right|^{2}}{\sum_{i=1}^{2} \mathbf{h}_{i}^{H}(\omega) \mathbf{\Gamma}_{\mathbf{v}}(\omega) \mathbf{h}_{i}(\omega)}.$$
(15)

In the particular case of $\mathbf{h}_1(\omega) = \mathbf{i}_i$ and $\mathbf{h}_2(\omega) = \mathbf{i}_j$, the binaural output SNR is equal to the input SNR, in which \mathbf{i}_i and \mathbf{i}_j are, respectively, the *i*th and *j*th columns of \mathbf{I}_{2M} (i.e., the $2M \times 2M$ identity matrix). According to (14) and (15), the binaural SNR gain can be expressed as

$$\mathcal{G}[\mathbf{h}_{1}(\omega), \mathbf{h}_{2}(\omega)] = \frac{\mathrm{oSNR}[\mathbf{h}_{1}(\omega), \mathbf{h}_{2}(\omega)]}{\mathrm{iSNR}(\omega)}$$

$$= \frac{\sum_{i=1}^{2} \left| \mathbf{h}_{i}^{H}(\omega) \mathbf{d}(\omega, 0) \right|^{2}}{\sum_{i=1}^{2} \mathbf{h}_{i}^{H}(\omega) \Gamma_{\mathbf{v}}(\omega) \mathbf{h}_{i}(\omega)}.$$
(16)

From the above definition, the following two measures that are very helpful for binaural fixed beamforming can be deduced:

• the binaural white noise gain (WNG):

$$\mathcal{W}[\mathbf{h}_{1}(\omega), \mathbf{h}_{2}(\omega)] = \frac{\sum_{i=1}^{2} \left| \mathbf{h}_{i}^{H}(\omega) \mathbf{d}(\omega, 0) \right|^{2}}{\sum_{i=1}^{2} \mathbf{h}_{i}^{H}(\omega) \mathbf{h}_{i}(\omega)}$$
(17)

• and the binaural DF:

$$\mathcal{D}[\mathbf{h}_{1}(\omega), \mathbf{h}_{2}(\omega)] = \frac{\sum_{i=1}^{2} \left| \mathbf{h}_{i}^{H}(\omega) \mathbf{d}(\omega, 0) \right|^{2}}{\sum_{i=1}^{2} \mathbf{h}_{i}^{H}(\omega) \Gamma_{\mathbf{d}}(\omega) \mathbf{h}_{i}(\omega)},$$
(18)

where $\Gamma_{d}(\omega)$ is defined in the previous section.

The beampattern is another fundamental performance measure for fixed beamformers. The binaural beampattern can be defined as

$$|\mathcal{B}[\mathbf{h}_1(\omega), \mathbf{h}_2(\omega), \theta]|^2 = \frac{1}{2} \sum_{i=1}^2 \left| \mathbf{h}_i^H(\omega) \mathbf{d}(\omega, \theta) \right|^2.$$
(19)

In order to have two meaningful estimates of the desired signal, we are going to extensively exploit the interaural coherence (IC) of the noise. It is well known that, in a multi-source environment, the IC (or its modulus) is important for source localization since it is very strongly related to the two principal binaural cues, i.e., the interaural time difference (ITD) and interaural level difference (ILD), that the brain uses to localize sounds. Psychoacoustically, the localization performance decreases when the IC decreases [33]. Furthermore, the IC affects significantly the perception for acoustic field width.

Let $A(\omega)$ and $B(\omega)$ be two zero-mean complex-valued random variables. The coherence function (CF) between $A(\omega)$ and $B(\omega)$ is defined as

$$\gamma_{AB}(\omega) = \frac{E[A(\omega)B^*(\omega)]}{\sqrt{E[|A(\omega)|^2]E[|B(\omega)|^2]}},$$
(20)

where the superscript * is the complex-conjugate operator. It is clear that $0 \le |\gamma_{AB}(\omega)|^2 \le 1$. For any pair of sensors (i, j), the input IC of the noise is simply the CF between $V_i(\omega)$ and $V_j(\omega)$, i.e.,

$$\gamma_{V_{i}V_{j}}(\omega) = \frac{E\left[V_{i}(\omega)V_{j}^{*}(\omega)\right]}{\sqrt{E\left[|V_{i}(\omega)|^{2}\right]E\left[|V_{j}(\omega)|^{2}\right]}}$$

$$= \frac{\mathbf{i}_{i}^{T}\mathbf{\Phi}_{\mathbf{v}}(\omega)\mathbf{i}_{j}}{\sqrt{\mathbf{i}_{i}^{T}\mathbf{\Phi}_{\mathbf{v}}(\omega)\mathbf{i}_{j} \times \mathbf{i}_{j}^{T}\mathbf{\Phi}_{\mathbf{v}}(\omega)\mathbf{i}_{j}}}$$

$$= \frac{\mathbf{i}_{i}^{T}\Gamma_{\mathbf{v}}(\omega)\mathbf{i}_{j}}{\sqrt{\mathbf{i}_{i}^{T}\Gamma_{\mathbf{v}}(\omega)\mathbf{i}_{j} \times \mathbf{i}_{j}^{T}\Gamma_{\mathbf{v}}(\omega)\mathbf{i}_{j}}}$$

$$= \gamma\left[\mathbf{i}_{i}(\omega),\mathbf{i}_{j}(\omega)\right].$$

$$(21)$$

For white noise, the input IC is $\gamma_w(\omega) = 0$, obviously. For diffuse noise, $\Gamma_v(\omega) = \Gamma_d(\omega)$, the input IC is $[\Gamma_d(\omega)]_{ii}$.

Similarly, we can define the output IC of the noise as the CF between the filtered noises in $Z_1(\omega)$ and $Z_2(\omega)$, i.e.,

$$\gamma[\mathbf{h}_{1}(\omega), \mathbf{h}_{2}(\omega)] = \frac{\mathbf{h}_{1}^{H}(\omega) \mathbf{\Phi}_{\mathbf{v}}(\omega) \mathbf{h}_{2}(\omega)}{\sqrt{\mathbf{h}_{1}^{H}(\omega) \mathbf{\Phi}_{\mathbf{v}}(\omega) \mathbf{h}_{1}(\omega) \times \mathbf{h}_{2}^{H}(\omega) \mathbf{\Phi}_{\mathbf{v}}(\omega) \mathbf{h}_{2}(\omega)}}$$
$$= \frac{\mathbf{h}_{1}^{H}(\omega) \Gamma_{\mathbf{v}}(\omega) \mathbf{h}_{2}(\omega)}{\sqrt{\mathbf{h}_{1}^{H}(\omega) \Gamma_{\mathbf{v}}(\omega) \mathbf{h}_{1}(\omega) \times \mathbf{h}_{2}^{H}(\omega) \Gamma_{\mathbf{v}}(\omega) \mathbf{h}_{2}(\omega)}}.$$
(22)

$$\gamma_{\rm w}[\mathbf{h}_1(\omega), \mathbf{h}_2(\omega)] = \frac{\mathbf{h}_1^H(\omega)\mathbf{h}_2(\omega)}{\sqrt{\mathbf{h}_1^H(\omega)\mathbf{h}_1(\omega) \times \mathbf{h}_2^H(\omega)\mathbf{h}_2(\omega)}}$$
(23)

and

$$\gamma_{d}[\mathbf{h}_{1}(\omega),\mathbf{h}_{2}(\omega)] = \frac{\mathbf{h}_{1}^{H}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{h}_{2}(\omega)}{\sqrt{\mathbf{h}_{1}^{H}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{h}_{1}(\omega) \times \mathbf{h}_{2}^{H}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{h}_{2}(\omega)}}.$$
(24)

In many approaches in the literature, the two derived filters $\mathbf{h}_1(\omega)$ and $\mathbf{h}_2(\omega)$ are collinear, i.e.,

$$\mathbf{h}_1(\omega) = \zeta(\omega)\mathbf{h}_2(\omega),\tag{25}$$

where $\zeta(\omega) \neq 0$ is a complex-valued number. In this case, one can check that $|\gamma[\mathbf{h}_1(\omega), \mathbf{h}_2(\omega)]| = |\gamma_w[\mathbf{h}_1(\omega), \mathbf{h}_2(\omega)]| = |\gamma_d[\mathbf{h}_1(\omega), \mathbf{h}_2(\omega)]| = 1$. Since the desired source signal is also fully coherent at any pair of sensors, both the desired signal and noise are perceived in the same region. As a result, our brain will have difficulties to separate them with binaural presentation and intelligibility will certainly be affected. For better separation between white noise and desired source, we should find orthogonal filters since, in this scenario, the output IC for white noise will be equal to 0, the same way is its corresponding input.

5. Binaural Heterophasic Superdirective Beamformer

In this section, we consider orthogonal binaural filters, i.e., $\mathbf{h}_1^H(\omega)\mathbf{h}_2(\omega) = 0$, since we want the output IC for white noise to be zero. We also want to maximize $\gamma_d[\mathbf{h}_1(\omega), \mathbf{h}_2(\omega)]$ so that not only the signals of interest from a point source at the two binaural outputs are coherent, the diffuse (or any correlated) noise at the binaural outputs will be perceived as less diffuse as possible. For that, we will exploit the maximum modes of this CF.

The symmetric matrix $\Gamma_{d}(\omega)$ can be diagonalized as [54]

$$\mathbf{U}^{T}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{U}(\omega) = \mathbf{\Lambda}(\omega), \tag{26}$$

where

$$\mathbf{U}(\omega) = \begin{bmatrix} \mathbf{u}_1(\omega) & \mathbf{u}_2(\omega) & \cdots & \mathbf{u}_{2M}(\omega) \end{bmatrix}$$
(27)

is an orthogonal matrix (note that each eigenvector is normalized according to the sign of its first value in our application because each eigenvector of $\Gamma_d(\omega)$ may have two opposite directions), i.e.,

$$\mathbf{U}^{T}(\omega)\mathbf{U}(\omega) = \mathbf{U}(\omega)\mathbf{U}^{T}(\omega) = \mathbf{I}_{2M}$$
(28)

and

$$\Lambda(\omega) = \operatorname{diag}[\lambda_1(\omega), \lambda_2(\omega), \dots, \lambda_{2M}(\omega)]$$
⁽²⁹⁾

is a diagonal matrix. The orthonormal vectors $\mathbf{u}_1(\omega), \mathbf{u}_2(\omega), \dots, \mathbf{u}_{2M}(\omega)$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_1(\omega), \lambda_2(\omega), \dots, \lambda_{2M}(\omega)$ of the matrix $\Gamma_d(\omega)$, where $\lambda_1(\omega) \ge \lambda_2(\omega) \ge \dots \ge \lambda_{2M}(\omega) > 0$.

It can be shown that the two orthogonal filters that maximize (24) are [55]

$$\mathbf{h}_{1}(\omega) = \frac{\mathbf{u}_{1}(\omega) + \mathbf{u}_{2M}(\omega)}{\sqrt{2}} = \mathbf{q}_{+,1}(\omega)$$
$$\mathbf{h}_{2}(\omega) = \frac{\mathbf{u}_{1}(\omega) - \mathbf{u}_{2M}(\omega)}{\sqrt{2}} = \mathbf{q}_{-,1}(\omega)$$
(30)

In this case, we get the first maximum mode of the CF:

$$\gamma_{\rm d}[\mathbf{q}_{+,1}(\omega), \mathbf{q}_{-,1}(\omega)] = \lambda_{\mp,1}(\omega), \tag{31}$$

with corresponding vectors $\mathbf{q}_{+,1}(\omega)$ and $\mathbf{q}_{-,1}(\omega)$, where

$$\lambda_{\mp,1}(\omega) = \frac{\lambda_1(\omega) - \lambda_{2M}(\omega)}{\lambda_1(\omega) + \lambda_{2M}(\omega)}$$

$$= \frac{\lambda_{-,1}(\omega)}{\lambda_{+,1}(\omega)}.$$
(32)

Similarly, we find that all the *M* maximum modes of the CF are

$$\gamma_{\rm d}[\mathbf{q}_{+,m}(\omega),\mathbf{q}_{-,m}(\omega)] = \lambda_{\mp,m}(\omega),\tag{33}$$

for m = 1, 2, ..., M, where

$$\lambda_{\mp,m}(\omega) = \frac{\lambda_m(\omega) - \lambda_{2M-m+1}(\omega)}{\lambda_m(\omega) + \lambda_{2M-m+1}(\omega)}$$

$$= \frac{\lambda_{-,m}(\omega)}{\lambda_{+,m}(\omega)}$$
(34)

and

$$\begin{cases} \mathbf{q}_{+,m}(\omega) = \frac{\mathbf{u}_{m}(\omega) + \mathbf{u}_{2M-m+1}(\omega)}{\sqrt{2}} \\ \mathbf{q}_{-,m}(\omega) = \frac{\mathbf{u}_{m}(\omega) - \mathbf{u}_{2M-m+1}(\omega)}{\sqrt{2}} \end{cases}$$
(35)

It can be verified that

$$\lambda_{\mp,1}(\omega) \ge \lambda_{\mp,2}(\omega) \ge \dots \ge \lambda_{\mp,M}(\omega). \tag{36}$$

From (35), the two semi-orthogonal matrices ($2M \times M$) can be written as

$$\mathbf{Q}_{+}(\omega) = \begin{bmatrix} \mathbf{q}_{+,1}(\omega) & \mathbf{q}_{+,2}(\omega) & \cdots & \mathbf{q}_{+,M}(\omega) \end{bmatrix},$$
(37)

$$\mathbf{Q}_{-}(\omega) = \begin{bmatrix} \mathbf{q}_{-,1}(\omega) & \mathbf{q}_{-,2}(\omega) & \cdots & \mathbf{q}_{-,M}(\omega) \end{bmatrix}, \quad (38)$$

where

$$\mathbf{Q}_{+}^{T}(\omega)\mathbf{Q}_{+}(\omega) = \mathbf{Q}_{-}^{T}(\omega)\mathbf{Q}_{-}(\omega) = \mathbf{I}_{M}, \qquad (39)$$
$$\mathbf{Q}_{+}^{T}(\omega)\mathbf{Q}_{-}(\omega) = \mathbf{Q}_{-}^{T}(\omega)\mathbf{Q}_{+}(\omega) = \mathbf{0}, \qquad (40)$$

$$\mathbf{Q}_{+}^{I}(\omega)\mathbf{Q}_{-}(\omega) = \mathbf{Q}_{-}^{I}(\omega)\mathbf{Q}_{+}(\omega) = \mathbf{0},$$
(40)

with \mathbf{I}_M being the $M\times M$ identity matrix. It can be shown that

$$\mathbf{Q}_{+}^{T}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{Q}_{-}(\omega) = \mathbf{Q}_{-}^{T}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{Q}_{+}(\omega)$$
$$= \mathbf{\Lambda}_{-}(\omega),$$
(41)

$$\mathbf{Q}_{+}^{T}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{Q}_{+}(\omega) = \mathbf{Q}_{-}^{T}(\omega)\mathbf{\Gamma}_{d}(\omega)\mathbf{Q}_{-}(\omega)$$

= $\mathbf{\Lambda}_{+}(\omega)$, (42)

where

$$\Lambda_{-}(\omega) = \frac{1}{2} \operatorname{diag}[\lambda_{-,1}(\omega), \lambda_{-,2}(\omega), \dots, \lambda_{-,M}(\omega)], \qquad (43)$$

$$\Lambda_{+}(\omega) = \frac{1}{2} \operatorname{diag}[\lambda_{+,1}(\omega), \lambda_{+,2}(\omega), \dots, \lambda_{+,M}(\omega)]$$
(44)

are two diagonal matrices of size $M \times M$.

Let *N* be a positive integer number with $2 \le N \le M$ (a different value of *N* gives a different degree of tradeoff between WNG and DF). We define the two semi-orthogonal matrices (of size $2M \times N$):

$$\mathbf{Q}_{+,N}(\omega) = \begin{bmatrix} \mathbf{q}_{+,1}(\omega) & \mathbf{q}_{+,2}(\omega) & \cdots & \mathbf{q}_{+,N}(\omega) \end{bmatrix},$$
(45)

$$\mathbf{Q}_{-,:N}(\omega) = \begin{bmatrix} \mathbf{q}_{-,1}(\omega) & \mathbf{q}_{-,2}(\omega) & \cdots & \mathbf{q}_{-,N}(\omega) \end{bmatrix}.$$
(46)

In the rest, we consider orthogonal filters of the forms:

$$\begin{cases} \mathbf{h}_{1}(\omega) = \mathbf{Q}_{+,:N}(\omega)\overline{\mathbf{h}}_{:N}(\omega) \\ \mathbf{h}_{2}(\omega) = \mathbf{Q}_{-,:N}(\omega)\overline{\mathbf{h}}_{:N}(\omega) \end{cases}$$
(47)

where

$$\overline{\mathbf{h}}_{:N}(\omega) = \begin{bmatrix} \overline{H}_{1}(\omega) & \overline{H}_{2}(\omega) & \cdots & \overline{H}_{N}(\omega) \end{bmatrix} \neq \mathbf{0}$$
(48)

is a complex-valued filter of length *N*. For this class of orthogonal filters, the output IC for diffuse noise is

$$\gamma_{d}[\mathbf{h}_{1}(\omega), \mathbf{h}_{2}(\omega)] = \frac{\overline{\mathbf{h}}_{:N}^{H}(\omega) \mathbf{\Lambda}_{-,N}(\omega) \overline{\mathbf{h}}_{:N}(\omega)}{\overline{\mathbf{h}}_{:N}^{H}(\omega) \mathbf{\Lambda}_{+,N}(\omega) \overline{\mathbf{h}}_{:N}(\omega)}$$
$$= \gamma_{d} \Big[\overline{\mathbf{h}}_{:N}(\omega) \Big],$$
(49)

where

$$\Lambda_{-,N}(\omega) = \frac{1}{2} \operatorname{diag}[\lambda_{-,1}(\omega), \lambda_{-,2}(\omega), \dots, \lambda_{-,N}(\omega)],$$
(50)

$$\mathbf{\Lambda}_{+,N}(\omega) = \frac{1}{2} \operatorname{diag}[\lambda_{+,1}(\omega), \lambda_{+,2}(\omega), \dots, \lambda_{+,N}(\omega)].$$
(51)

It can be shown that

$$1 \ge \gamma \left[\overline{\mathbf{h}}_{:1}(\omega) \right] \ge \gamma \left[\overline{\mathbf{h}}_{:2}(\omega) \right] \ge \dots \ge \gamma \left[\overline{\mathbf{h}}_{:M}(\omega) \right] \ge 0.$$
(52)

With (47), the binaural WNG, DF, and power beampattern can be expressed as, respectively,

$$\mathcal{W}\left[\overline{\mathbf{h}}_{:N}(\omega)\right] = \frac{\overline{\mathbf{h}}_{:N}^{H}(\omega)\mathbf{C}(\omega,0)\mathbf{C}^{H}(\omega,0)\overline{\mathbf{h}}_{:N}(\omega)}{2\overline{\mathbf{h}}_{:N}^{H}(\omega)\overline{\mathbf{h}}_{:N}(\omega)},$$
(53)

$$\mathcal{D}\left[\overline{\mathbf{h}}_{:N}(\omega)\right] = \frac{\overline{\mathbf{h}}_{:N}^{H}(\omega)\mathbf{C}(\omega,0)\mathbf{C}^{H}(\omega,0)\overline{\mathbf{h}}_{:N}(\omega)}{2\overline{\mathbf{h}}_{:N}^{H}(\omega)\mathbf{\Lambda}_{+,N}(\omega)\overline{\mathbf{h}}_{:N}(\omega)},$$
(54)

and

$$\left|\mathcal{B}\left[\overline{\mathbf{h}}_{:N}(\omega),\theta\right]\right|^{2} = \frac{\overline{\mathbf{h}}_{:N}^{H}(\omega)\mathbf{C}(\omega,\theta)\mathbf{C}^{H}(\omega,\theta)\overline{\mathbf{h}}_{:N}(\omega)}{2},\tag{55}$$

where

$$\mathbf{C}(\omega,\theta) = \begin{bmatrix} \mathbf{Q}_{+,:N}^{T}(\omega)\mathbf{d}(\omega,\theta) & \mathbf{Q}_{-,:N}^{T}(\omega)\mathbf{d}(\omega,\theta) \end{bmatrix}$$
(56)

is a matrix of size $N \times 2$ with the distortionless constraint being

$$\mathbf{C}^{H}(\omega,0)\overline{\mathbf{h}}_{:N}(\omega) = \mathbf{1} = \begin{bmatrix} 1\\1 \end{bmatrix}.$$
(57)

To fulfill this constraint, we must take $N \ge 2$.

Substituting (47) into (12) and using the distortionless constraint, the variance of $Z_i(\omega)$ becomes

$$\phi_{Z_i}(\omega) = \phi_X(\omega) + \phi_{V_1}(\omega) \times \overline{\mathbf{h}}_{:N}^H(\omega) \mathbf{Q}_{\pm,:N}^T(\omega) \Gamma_{\mathbf{v}}(\omega) \mathbf{Q}_{\pm,:N}(\omega) \overline{\mathbf{h}}_{:N}(\omega),$$
(58)

where $\mathbf{Q}_{\pm,:N}(\omega) = \mathbf{Q}_{+,:N}(\omega)$ for $\phi_{Z_1}(\omega)$ and $\mathbf{Q}_{\pm,:N}(\omega) = \mathbf{Q}_{-,:N}(\omega)$ for $\phi_{Z_2}(\omega)$. In the case of diffuse-plus-white noise, i.e., $\Gamma_{\mathbf{v}}(\omega) = \Gamma_{\mathbf{d}}(\omega) + \alpha \mathbf{I}_{2M}$, where α is a parameter that determines the relative level between the diffuse and white noises, (58) simplifies to

$$\phi_{Z_i}(\omega) = \phi_X(\omega) + \phi_{V_1}(\omega) \times \left[\overline{\mathbf{h}}_{:N}^H(\omega) \mathbf{\Lambda}_{+,N}(\omega) \overline{\mathbf{h}}_{:N}(\omega) + \alpha \overline{\mathbf{h}}_{:N}^H(\omega) \overline{\mathbf{h}}_{:N}(\omega)\right],$$
(59)

showing that $\phi_{Z_1}(\omega) = \phi_{Z_2}(\omega)$. Again, using (57), we find that the cross-correlation between $Z_1(\omega)$ and $Z_2(\omega)$ is

$$\begin{aligned} \phi_{Z_1 Z_2}(\omega) &= E[Z_1(\omega) Z_2^*(\omega)] \\ &= \phi_X(\omega) + \phi_{V_1}(\omega) \overline{\mathbf{h}}_{:N}^H(\omega) \mathbf{Q}_{+,:N}^T(\omega) \mathbf{\Gamma}_{\mathbf{v}}(\omega) \mathbf{Q}_{-,:N}(\omega) \overline{\mathbf{h}}_{:N}(\omega), \end{aligned}$$
(60)

whose form for diffuse-plus-white noise is

$$\phi_{Z_1Z_2}(\omega) = \phi_X(\omega) + \phi_{V_1}(\omega)\overline{\mathbf{h}}_{:N}^H(\omega)\mathbf{\Lambda}_{-,N}(\omega)\overline{\mathbf{h}}_{:N}(\omega), \tag{61}$$

which, as expected, does not depend on the white noise. For $\Gamma_v(\omega) = \Gamma_d(\omega) + \alpha I_{2M}$, the output IC of the estimated signals is

$$\gamma_{Z_1 Z_2}(\omega) = \frac{\phi_{Z_1 Z_2}(\omega)}{\sqrt{\phi_{Z_1}(\omega)\phi_{Z_2}(\omega)}}$$
$$= \frac{\mathrm{iSNR}(\omega) + \overline{\mathbf{h}}_{:N}^H(\omega)\mathbf{\Lambda}_{-,N}(\omega)\overline{\mathbf{h}}_{:N}(\omega)}{\mathrm{iSNR}(\omega) + \overline{\mathbf{h}}_{:N}^H(\omega)\mathbf{\Lambda}_{+,N}(\omega)\overline{\mathbf{h}}_{:N}(\omega) + \alpha\overline{\mathbf{h}}_{:N}^H(\omega)\overline{\mathbf{h}}_{:N}(\omega)}.$$
(62)

We deduce that for large input SNRs, the localization cues of the estimated signals depend mostly on the ones of the desired signal, while for low SNRs, they depend mostly on the ones of the diffuse-plus-white noise.

One possible binaural superdirective beamformer can be obtained by minimizing the sum of the filtered diffuse noise signals subject to the distortionless constraint in (57), i.e.,

ы

$$\min_{\overline{\mathbf{h}}_{:N}(\omega)} 2\overline{\mathbf{h}}_{:N}^{II}(\omega) \mathbf{\Lambda}_{+,N}(\omega) \overline{\mathbf{h}}_{:N}(\omega)$$
s. t. $\overline{\mathbf{h}}_{:N}^{H}(\omega) \mathbf{C}(\omega, 0) = \mathbf{1}^{T}$. (63)

We easily get

$$\overline{\mathbf{h}}_{:N,\text{BSD}}(\omega) = \mathbf{\Lambda}_{+,N}^{-1}(\omega)\mathbf{C}(\omega,0) \left[\mathbf{C}^{H}(\omega,0)\mathbf{\Lambda}_{+,N}^{-1}(\omega)\mathbf{C}(\omega,0)\right]^{-1}\mathbf{1}$$
(64)

and the corresponding binaural DF is

$$\mathcal{D}\left[\overline{\mathbf{h}}_{:N,\text{BSD}}(\omega)\right] = \frac{1}{\mathbf{1}^{T} \left[\mathbf{C}^{H}(\omega,0)\mathbf{\Lambda}_{+,N}^{-1}(\omega)\mathbf{C}(\omega,0)\right]^{-1}\mathbf{1}}.$$
(65)

Therefore, the proposed binaural superdirective beamformer is

$$\begin{cases} \mathbf{h}_{1,\text{BSD}}(\omega) = \mathbf{Q}_{+,:N}(\omega)\overline{\mathbf{h}}_{:N,\text{BSD}}(\omega) \\ \mathbf{h}_{2,\text{BSD}}(\omega) = \mathbf{Q}_{-,:N}(\omega)\overline{\mathbf{h}}_{:N,\text{BSD}}(\omega) \end{cases}$$
(66)

Note that another form of the binaural superdirective beamformer can be derived by maximizing the binaural DF in (18) subject to the distortionless constraints, which will be left to the reader's investigation to make the paper concise.

6. Experiments and Analysis

In this section, we study the performance of the developed binaural heterophasic superdirective beamforming method and compare it to the monaural superdirective beamformer through experiments. For fair comparison, the orders of the binaural heterophasic and the monaural superdirective beamformers are set to the same number so the DF of the two beamformers would be similar.

6.1. Performance Analysis

We first evaluate the beampattern (given in (55)) of the binaural superdirective beamformer. A ULA is used with an interelement spacing equal to 1 cm. The beampatterns of the derived beamformer are plotted in Figure 3, where M = 6, 8, 10, 12, and N = M/2, at f = 1 kHz. Note that, given a ULA with M microphone sensors, one can design binaural superdirective beamformers of order from 1 to M/2. When the order increases, the DF becomes larger but the WNG becomes smaller. In this work, we only show the case with N = M/2 in the simulation for maximum DF. Two-dimensional (2D) plots of the corresponding beampatterns are shown in Figure 4; one can observe that, in all cases, they are almost frequency invariant.

Next, We study the performance of the binaural heterophasic superdirective beamformer in terms of WNG and DF, according to (53) and (54), respectively. The results for the WNG and DF are plotted in Figure 5, where the first-, second-, third-, and fourth-order binaural superdirective beamformers are designed with M = 6, 8, 10, 12, respectively (this is the basic requirement for the design of the binaural heterophasic superdirective beamformer as shown in Section 5). One can see that the WNGs of binaural superdirective beamformer decrease with the order while the DFs of beamformers increase with the order. Besides, for each order, the DF does not change much with frequency, which is an important property for processing broadband signals like speech. Figure 6 plots the WNGs and DFs of the binaural superdirective beamformers versus parameter N, where $2 \le N \le M/2$. This parameter is introduced to gain flexibility for achieving compromise between WNG



and DF. As seen, the DF increases while the WNG decreases with the increase of the value of *N*. In practice, one can tune the parameter *N* according to the application requirement.

Figure 3. Beampatterns of the binaural heterophasic superdirective beamformer with various numbers of microphones: (a) M = 6, (b) M = 8, (c) M = 10, and (d) M = 12. Conditions of simulation: f = 1 kHz, $\delta = 1 \text{ cm}$, and N = M/2.

Next, we study the ICs of the binaural heterophasic and conventional superdirective beamformers under the same conditions according to (23) and (24). Figure 7 plots the output ICs of both beamformers as a function of frequency in white and diffuse noises, respectively. As seen, in the diffuse noise case, the ICs of both beamformers are equal to one within the studied frequency range; in the white noise case, the IC of the binaural superdirective beamformer is equal to zero while for the conventional superdirective beamformer it is equal to one. This means that in the two output signals of the binaural superdirective beamformer, the speech signal is completely coherent, while the white noise is completely incoherent; so, the output signals correspond to the heterophasic case as discussed in Section 1, in which the speech and white noise can be regarded as two separate direction sources in space.

Figure 8 plots the IC magnitude of the outputs of the binaural heterophasic superdirective beamformer, which is given in (62), versus frequency in different input SNR conditions. One can see that this IC increases with frequency. This is due to the fact that white noise amplification mainly happens at low frequencies. The output IC in white noise is zero, causing the low-frequency output IC of the entire signal to approach zero. This shows the impact of white noise amplification from the perspective of the output IC. For a fixed frequency, it is seen that the output IC increases with the input SNR, and it approaches one at a high input SNR. This can be easily explained: as the input SNR increases, the desired signal component dominates the beamforming output. The output IC of the desired signal is one, so the output IC of the two output signals in this condition also approaches one. Consequently, one can conclude that, for high input SNRs, the localization cues of the estimated signals depend mostly on the ones of the desired signal, while at low SNRs, they depend mostly on the ones of the noise.



Figure 4. Beampatterns of the binaural heterophasic superdirective beamformer as a function of the frequency with different numbers of microphones: (a) M = 6, (b) M = 8, (c) M = 10, and (d) M = 12. Conditions of simulation: $\delta = 1.0$ cm and N = M/2.



Figure 5. WNGs and DFs of the 1st-, 2nd-, 3rd-, and 4th-order binaural heterophasic superdirective beamformers versus frequency: (a) WNGs and (b) DFs. The sensor spacing: $\delta = 1$ cm.



Figure 6. WNGs and DFs of the 1st-, 2nd-, 3rd-, and 4th-order binaural heterophasic superdirective beamformers versus parameter *N*: (a) WNGs and (b) DFs. Conditions of simulation: f = 1 kHz and $\delta = 1$ cm.



Figure 7. Magnitude of the output ICs: (a) diffuse noise and (b) white noise. Conditions of simulation: M = 8 for the binaural superdirective beamformer, M = 3 for the conventional superdirective beamformer, and $\delta = 1$ cm. Note that the output IC of the binaural superdirective beamformer in diffuse noise is frequency-dependent; it approaches 1 under 4 kHz but decreases as the frequency increases.



Figure 8. Output IC of the estimated signals (see (62)) as a function of frequency and input SNR. Conditions of simulation: M = 8, N = 4, $\alpha = 1$, and $\delta = 1$ cm.

6.2. Experiments in Real Environments

In this subsection, we evaluate the performance of the proposed binaural heterophasic superdirective beamformer in real acoustic environments. The experiments were conducted in a $10.5 \times 6 \times 2.8$ m conference room. A ULA is used, which consists of 8 microphones, where the elements spacing is $\delta = 1.1$ cm. The SNR of the microphones is 60 dB(A). A photo of the designed array and the experimental setup are shown in Figure 9. To make the experiments repeatable, we first used the microphone array to record sound signals from a loudspeaker located in the ULA's endfire direction. Then, both the conventional and binaural superdirective beamformers were implemented to get the outputs.

Figure 10 plots the time-domain signals and their spectrograms of the output signals of the conventional and binaural superdirective beamformers. It is clearly seen that the outputs of both beamformers suffer from serious white noise amplification, where the desired signal is almost covered by the white noise. It is also seen from spectrograms that white noise amplification mainly occurs at low frequencies.

As emphasized previously, the main advantage of the proposed binaural superdirective beamforming method is to have the human binaural auditory system to better separate the signal of interest from white noise after beamforming. To confirm this, we performed some subjective listening experiments. Firstly, we obtained a series of output signals of the implemented conventional and binaural superdirective beamformers. Specifically, we extracted seven audio clips from the "Voice of America" with each of length of 20 s. After playing and recording through the loudspeaker and the microphone array shown in Figure 9, we use the two kind of beamformers to perform superdirective beamforming to obtain the output signals. Then, five subjects were asked to listen to the output signals and draw up the zones of the sound source and white noise on the horizontal plane within a predesigned circle. Finally, we extracted the images sketched by each subject and averaged them to get the experimental results (note that in these experiments, we selected five subjects from the CIAIC-Center of Intelligent Acoustics and Immersive Communications, who are experienced in acoustic analysis and can clearly distinguish auditory events). Here we only provide four different zones, i.e., front, back, left, and right sides, to ask the subjects to choose, where an illustration of the test is shown in Figure 11.



(c)

Figure 9. A photo of the designed array and the experimental setup for evaluating the binaural and conventional superdirective beamformers: (**a**) photo of the designed eight-microphone array, (**b**) a close view photo of the experimental setup, and (**c**) a wide angle photo of the experimental setup.

Figure 12 presents the auditory maps averaged from the five subjects from the horizontal plane. As can be seen, for the conventional superdirective beamformer, all signals (desired signal plus white noise) are perceived to be in the middle of the head, which corresponds to the homophasic case. Oppositely, for the binaural superdirective beamforming, the signal of interest is perceived to be in the median plane of the head while white noise is located in each side of the ear, which corresponds to the heterophasic case. As discussed previously and summarized in Table 1, the speech intelligibility in heterophasic case is higher than the homophasic case (approximately 4 dB higher [49]). Consequently, the proposed binaural superdirective beamformer has better intelligibility than the conventional one.



Figure 10. Monaural and binaural superdirective beamformers in a conference room: (a) output of the 2nd-order monaural superdirective beamformer and its spectrogram with M = 3 and $\delta = 1.1$ cm, and (b) output of the 2nd-order binaural superdirective beamformer and its spectrogram with M = 8 and $\delta = 1.1$ cm.



Figure 11. An illustration of the auditory map for subjects (horizontal-plane). During the test, the subjects were asked to mark the areas according to the sound source location they heard through headphones.



Figure 12. The average auditory map marked by the listening subjects: (a) monaural superdirective beamformer and (b) binaural superdirective beamformer. The blue waves refer to the region in which the desired speech is heard, and the region filled with disorderly dots refer to the region in which the white noise is heard. Conditions of experiment: M = 8, N = 4, and $\delta = 1.1$ cm.

7. Conclusions

In this paper, we addressed the problem of superdirective beamforming with smallspacing microphone arrays. While it can achieve the maximum spatial gain to suppress acoustic noise, the traditional superdirective beamformer suffers from white noise amplification, which is particularly serious at low frequencies. Many methods were developed in the literature to deal with this problem, but they all pay a price of sacrificing the DF and the resulting beamformers may no longer be superdirective. Motivated by studies and facts in psychoacoustics, we developed in this paper a binaural heterophasic superdirective beamformer, which consists of two sub-beamforming filters, each for one of the binaural channels. These two sub-beamformers are constrained to be orthogonal to each other to minimize the IC of the white noise components in the binaural outputs while maximize the IC of the diffuse noise components. As a result, the signal of interest in the binaural superdirective beamformer's outputs is in phase while the white noise is random in phase, so that the human auditory system is able to more easily separate the acoustic signal of interest from white noise by listening to the outputs of the proposed beamformer. Simulations and experiments were carried out to evaluate the performance of the proposed binaural superdirective beamformer. The results corroborate with the theoretical analysis and confirm that the binaural superdirective beamforming corresponds to the heterophasic case studied in psychoacoustics. Based on the listening tests shown in the psychoacoustic study, one can conclude that the improvement in intelligibility is expected to be 4 dB at low SNR conditions in accordance with the psychoacoustic experiments verified in the literature.

Author Contributions: Conceptualization, J.C. and J.B.; methodology, J.C. and J.B.; software, Y.W. and J.J.; validation, Y.W., J.J. and G.H.; formal analysis, Y.W., J.J. and G.H.; investigation, Y.W., J.J. and G.H.; writing—original draft preparation, J.C., J.B. and Y.W.; writing—review and editing, G.H., J.C. and J.B.; visualization, Y.W. and J.J.; supervision, J.C. and J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

References

- 1. Benesty, J.; Chen, J.; Pan, C. Fundamentals of Differential Beamforming; Springer: Berlin, Germany, 2016.
- 2. Benesty, J.; Chen, J.; Huang, Y. Microphone Array Signal Processing; Springer: Berlin, Germany, 2008.
- Elko, G.W. Superdirectional microphone arrays. In Acoustic Signal Processing for Telecommunication; Springer: Berlin, Germany, 2000; pp. 181–237.
- 4. Brandstein, M.; Ward, D. Microphone Arrays: Signal Processing Techniques and Applications; Springer: Berlin, Germany, 2001.

- Johnson, D.H.; Dudgeon, D.E. Array Signal Processing: Concepts and Techniques; PTR Prentice Hall: Englewood Cliffs, NJ, USA, 1993.
- 6. Huang, G.; Benesty, J.; Chen, J. On the Design of Frequency-Invariant Beampatterns with Uniform Circular Microphone Arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2017, 25, 1140–1153. [CrossRef]
- Huang, G.; Chen, J.; Benesty, J. Insights into Frequency-Invariant Beamforming with Concentric Circular Microphone Arrays. IEEE/ACM Trans. Audio Speech Lang. Process. 2018, 26, 2305–2318. [CrossRef]
- 8. Benesty, J.; Chen, J.; Cohen, I. Design of Circular Differential Microphone Arrays; Springer: Berlin, Germany, 2015.
- 9. Cox, H.; Zeskind, R.M.; Kooij, T. Practical supergain. IEEE Trans. Acoust. Speech Signal Process. 1986, 34, 393–398. [CrossRef]
- Huang, G.; Benesty, J.; Cohen, I.; Chen, J. A simple theory and new method of differential beamforming with uniform linear microphone arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2020, 28, 1079–1093. [CrossRef]
- Li, C.; Benesty, J.; Huang, G.; Chen, J. Subspace superdirective beamformers based on joint diagonalization. In Proceedings of the IEEE ICASSP, Shanghai, China, 20–25 March 2016; pp. 400–404.
- Pan, C.; Chen, J.; Benesty, J. Design of robust differential microphone arrays with orthogonal polynomials. J. Acoust. Soc. Am. 2015, 138, 1079–1089. [CrossRef]
- Huang, G.; Chen, J.; Benesty, J. Design of planar differential microphone arrays with fractional orders. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2019, 28, 116–130. [CrossRef]
- 14. Huang, G.; Chen, J.; Benesty, J. A flexible high directivity beamformer with spherical microphone arrays. *J. Acoust. Soc. Am.* **2018**, 143, 3024–3035. [CrossRef]
- Jin, J.; Chen, J.; Benesty, J.; Wang, Y.; Huang, G. Heterophasic Binaural Differential Beamforming for Speech Intelligibility Improvement. *IEEE Trans. Veh. Technol.* 2020, 69, 13497–13509
 [CrossRef]
- 16. Elko, G.W.; Meyer, J. Microphone arrays. In Springer Handbook of Speech Processing; Springer: Berlin, Germany, 2008; pp. 1021–1041.
- 17. Lotter, T.; Vary, P. Dual-channel speech enhancement by superdirective beamforming. *EURASIP J. Appl. Signal Process.* 2006, 1, 063297. [CrossRef]
- 18. Huang, G.; Benesty, J.; Chen, J. Design of robust concentric circular differential microphone arrays. *J. Acoust. Soc. Am.* 2017, 141, 3236–3249. [CrossRef]
- 19. Benesty, J.; Chen, J. Study and Design of Differential Microphone Arrays; Springer: Berlin, Germany, 2012.
- 20. Mabande, E.; Schad, A.; Kellermann, W. Design of robust superdirective beamformers as a convex optimization problem. In Proceedings of the IEEE ICASSP, Taipei, Taiwan, 19–24 April 2009; pp. 77–80.
- 21. Berkun, R.; Cohen, I.; Benesty, J. Combined beamformers for robust broadband regularized superdirective beamforming. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2015, 23, 877–886. [CrossRef]
- 22. Crocco, M.; Trucco, A. Design of robust superdirective arrays with a tunable tradeoff between directivity and frequency-invariance. *IEEE Trans. Signal Process.* 2011, 59, 2169–2181. [CrossRef]
- Huang, G.; Benesty, J.; Chen, J. Superdirective Beamforming Based on the Krylov Matrix. *IEEE/ACM Trans. Audio Speech Lang.* Process. 2016, 24, 2531–2543. [CrossRef]
- 24. Pan, C.; Chen, J.; Benesty, J. Reduced-order robust superdirective beamforming with uniform linear microphone arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2016, 24, 1548–1559. [CrossRef]
- Crocco, M.; Trucco, A. Stochastic and analytic optimization of sparse aperiodic arrays and broadband beamformers with robust superdirective patterns. *IEEE Trans. Audio, Speech, Lang. Process.* 2012, 20, 2433–2447. [CrossRef]
- Berkun, R.; Cohen, I.; Benesty, J. A tunable beamformer for robust superdirective beamforming. In Proceedings of the IEEE IWAENC, Xi'an, China , 13–16 September 2016; pp. 1–5.
- 27. Blauert, J. Spatial Hearing: The Psychophysics of Human Sound Localization; MIT Press: Cambridge, MA, USA, 1997.
- 28. Blauert, J.; Lindemann, W. Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J. Acoust. Soc. Am.* **1986**, *79*, 806–813. [CrossRef] [PubMed]
- 29. Hirsh, I.J. The Influence of Interaural Phase on Interaural Summation and Inhibition. J. Acoust. Soc. Am. 1948, 20, 536–544. [CrossRef]
- 30. Hirsh, I.J. The relation between localization and intelligibility. J. Acoust. Soc. Am. 1950, 22, 196–200. [CrossRef]
- 31. Jeffress, L.A.; Blodgett, H.C.; Deatherage, B.H. Effect of interaural correlation on the precision of centering a noise. *J. Acoust. Soc. Am.* **1962**, *34*, 1122–1123. [CrossRef]
- 32. Kollmeier, B.; Brand, T.; Meyer, B. Perception of speech and sound. In *Springer Handbook of Speech Processing*; Springer: Berlin, Germany, 2008; pp. 61–82.
- Zimmer, U.; Macaluso, E. High binaural coherence determines successful sound localization and increased activity in posterior auditory areas. *Neuron* 2005, 47, 893–905. [CrossRef]
- 34. Miller, G.A.; Licklider, J.C. The intelligibility of interrupted speech. J. Acoust. Soc. Am. 1950, 22, 167–173. [CrossRef]
- 35. Blauert, J. Sound localization in the median plane. Acta Acust. United Acust. 1969, 22, 205–213.
- 36. Jeffress, L.A. A place theory of sound localization. J. Comp. Physiol. Psychol. 1948, 41, 35. [CrossRef] [PubMed]
- 37. Sandel, T.; Teas, D.; Feddersen, W.; Jeffress, L. Localization of sound from single and paired sources. *J. Acoust. Soc. Am.* **1955**, 27, 842–852. [CrossRef]
- 38. Hirsh, I.J.; Pollack, I. The role of interaural phase in loudness. J. Acoust. Soc. Am. 1948, 20, 761–766. [CrossRef]

- 39. Kock, W. Binaural localization and masking. J. Acoust. Soc. Am. 1950, 22, 801–804. [CrossRef]
- 40. Jeffress, L.A.; Blodgett, H.C.; Sandel, T.T.; Wood, C.L, III. Masking of tonal signals. J. Acoust. Soc. Am. 1956, 28, 416–426. [CrossRef]
- Jeffress, L.A.; Robinson, D.E. Formulas for the coefficient of interaural correlation for noise. J. Acoust. Soc. Am. 1962, 34, 1658–1659. [CrossRef]
- 42. Licklider, J.C.; Miller, G.A. The Perception of Speech. In *Handbook of Experimental Psychology*; Stevens, S.S., Ed.; Wiley: Hoboken, NJ, USA, 1951; pp. 1040–1074.
- 43. Beranek, L.L. Acoustics; Acoustic Society of America: Woodbury, NY, USA, 1986.
- 44. Gerald, K.; Colburn, H.S. Informational Masking in Speech Recognition; Springer: Cham, Switzerland, 2017.
- 45. Zobel, B.H.; Wagner, A.; Sanders, L.D.; Ba_sskent, D. Spatial release from informational masking declines with age: Evidence from a detection task in a virtual separation paradigm. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, 146, 548–566. [CrossRef]
- 46. Moore, B.C. Effects of hearing loss and age on the binaural processing of temporal envelope and temporal fine structure information. *Hear. Res.* **2020**, 107991. [CrossRef]
- 47. Esther, S.; van de Par, S. The role of reliable interaural time difference cues in ambiguous binaural signals for the intelligibility of multitalker speech. J. Acoust. Soc. Am. 2020, 147, 4041–4054.
- 48. Jutras, B.; Lagacé, J.; Koravand, A. The development of auditory functions. In *Handbook of Clinical Neurology*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 143–155.
- 49. Licklider, J.C.R. The Influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Am.* **1948**, 20, 150–159. [CrossRef]
- 50. Moore, B.C. An Introduction to the Psychology of Hearing; Brill: Leiden, The Netherlands, 2012.
- 51. Marquardt, D.; Hohmann, V.; Doclo, S. Interaural coherence preservation in multi-channel Wiener filtering-based noise reduction for binaural hearing aids. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 2162–2176. [CrossRef]
- 52. Hadad, E.; Marquardt, D.; Doclo, S. Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2015, 23, 2449–2464. [CrossRef]
- 53. Cox, H.; Zeskind, R.; Owen, M. Robust adaptive beamforming. *IEEE Trans. Acoust. Speech Signal Process.* **1987**, *35*, 1365–1376. [CrossRef]
- 54. Golub, G.H.; Loan, C.F.V. Matrix Computations, 3rd ed.; The Johns Hopkins University Press: Baltimore, MD, USA, 1996.
- 55. Eaton, M.L. A maximization problem and its application to canonical correlation. J. Multivar. Anal. 1976, 6, 422–425. [CrossRef]