applied sciences

MDPI

Article

Improved Distributed Minimum Variance Distortionless Response (MVDR) Beamforming Method Based on a Local Average Consensus Algorithm for Bird Audio Enhancement in Wireless Acoustic Sensor Networks

Jiangjian Xie [1], Xingguang Li [1], Zhaoliang Xing [2], Bowen Zhang [1], Weidong Bao [3,*] and Junguo Zhang [1,*]

[1] School of Technology, Beijing Forestry University, Beijing 100083, China
[2] State Key Laboratory of Advanced Transmission Technology, Global Energy Interconnection Research Institute Co. Ltd., Beijing 102200, China
[3] School of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China
* Correspondence: wdbao@bjfu.edu.cn (W.B.); zhangjunguo@bjfu.edu.cn (J.Z.)

Received: 13 June 2019; Accepted: 29 July 2019; Published: 2 August 2019

Featured Application: With this bird audio enhancement method, the bird audio collected through the WASN (Wireless Acoustic Sensor Network) can be processed to produce better quality audio, which is more suitable for bird species identification based on the bird audio, then a higher accuracy of identification will be achieved.

Abstract: Currently, wireless acoustic sensor networks (WASN) are commonly used for wild bird monitoring. To better realize the automatic identification of birds during monitoring, the enhancement of bird audio is essential in nature. Currently, distributed beamformer is the most suitable method for bird audio enhancement of WASN. However, there are still several disadvantages of this method, such as large noise residue and slow convergence rate. To overcome these shortcomings, an improved distributed minimum variance distortionless response (IDMVDR) beamforming method for bird audio enhancement in WASN is proposed in this paper. In this method, the average metropolis weight local average consensus algorithm is first introduced to increase the consensus convergence rate, then a continuous spectrum update algorithm is proposed to estimate the noise power spectral density (PSD) to improve the noise reduction performance. Lastly, an MVDR beamformer is introduced to enhance the bird audio. Four different network topologies of the WASNs were considered, and the bird audio enhancement was performed on these WASNs to validate the effectiveness of the proposed method. Compared with two classical methods, the results show that the Segmental signal to noise ratio (SegSNR), mean square error (MSE), and perceptual evaluation of speech quality (PESQ) obtained by the proposed method are better and the consensus rate is faster, which means that the proposed method performs better in audio quality and convergence rate, and therefore it is suitable for WASN with dynamic topology.

Keywords: bird audio enhancement; wireless acoustic sensor networks; IDMVDR; local average consensus algorithm

1. Introduction

Bird species have been encountering increasing threats in recent years, which has attracted wide attention from ornithologists [1]. It is significant to monitor bird species for bird protection. With wireless acoustic sensor networks (WASN) becoming more and more popular, WASNs are commonly used for monitoring bird audio long-term [2]. Automatic bird species identification provides a suitable way to analyze the huge audio data from long-term monitoring programs [3]. However, the bird audio collected in nature is always accompanied by ambient noises, which consequently affect the accuracy of the bird species identification [4]. Therefore, audio enhancement should be carried out before identification, to improve identification accuracy.

By exploiting the spatial properties of speech and noise signals, WASN techniques can significantly outperform single-channel techniques in terms of improving interference suppression and reducing speech distortion [5–9]. Although WASN has many advantages, it also has several challenges, such as the limited energy and calculation ability of each node. There are two kinds of audio enhancement algorithms for WASNs. The first is centralized, all the data is transferred to a so-called fusion center (FC) for further enhancement. The second is distributed, the enhancements are performed on all the nodes, which means the computations are decomposed to all the nodes, then the amount of data transferred between nodes is reduced. Compared with the centralized method, the distributed method is typically preferred because of its lower energy consumption and higher scalability [10]. Several studies have been conducted on distributed methods of speech enhancement in WASNs. Three distributed audio enhancement algorithms [11–13] were presented in full-connect or tree topology networks, which can achieve the optimal estimated beamformer results at each node by using data compression to reduce energy consumption. Yuan and Hendriks proposed a distributed delay and sum beamformer (DDSB) for speech enhancement via randomized gossip in any topology of WASN—the same result as centralized beamformers can be obtained in each node of WASN [14]. Moreover, on the basis of previous studies, an improved general distributed synchronous averaging (IGDSA) algorithm was proposed [15] and the proposed IGDSA algorithm presented faster coverage rate than original synchronous communication scheme, especially for non-regular networks. Li proposed a speech enhancement algorithm through a distributed minimum variance distortionless response (DMVDR) beamformer. With this algorithm, each node in the WASN can obtain the same estimation result by communicating with neighboring nodes [16].

For any network topology, the distributed beamformer can estimate the desired signal at each node by only exchanging information with its neighbors. Only local signal interchanges are implemented in the distributed beamformer, which brings the advantages of higher robustness and scalability for sensor networks with a large node number and dynamic topology. So, the distributed beamformer is chosen for bird audio enhancement in this paper. Despite the rapid development of the speech enhancement technology of WASNs, research on the enhancement of bird audio using WASNs in forest areas is scarce. The noises in forest areas are always time-variant. However, in the traditional distributed speech enhancement of WASNs, the noise power spectral density (PSD) is commonly estimated by means of temporal averaging over noise-only segments [11–16], which may introduce large residual noise [17]. Moreover, the convergence rates of existing distributed beamformer methods are not fast enough, such as the random gossip algorithm of DDSB [14,15]. These restrictions obstruct the practical utilization of the distributed beamformer method. In this paper, an improved distributed minimum variance distortionless response (IDMVDR) beamforming method is proposed to realize the bird audio enhancement, in which the continuous spectrum update algorithm is introduced to estimate the noise PSD, and then the average metropolis weight algorithm is proposed to update each signal of all nodes to converge to the optimal solution of the centralized beamformer. At last, simulation experiments with bird audios were performed to validate the effectiveness of the proposed method.

This paper is organized as follows. In Section 2, firstly, the optimal centralized beamformer is discussed. Then, the average metropolis weight local average consensus algorithm is briefly reviewed. Subsequently, the distributed MVDR (DMVDR) beamformer is described in detail. Finally, the improved noise PSD algorithm is stated. After that, the iterative procedures of IDMVDR are listed. In Section 3, the simulation experiments, results, and analysis of the convergence rate and noise reduction performance are presented. Conclusions are drawn in Section 4.

2. Method Analysis

The DMVDR beamforming method uses a local average consensus algorithm to update the beamforming value in each node by only communicating with neighbor nodes. Each node can obtain the same estimated beamforming result as centralized MVDR [10,14,15]. Based on this method, through improving both the local average consensus algorithm and noise estimation algorithm, the IDMVDR beamforming algorithm was proposed to enhance the bird audios which are collected through the WASN in nature. The process flow chart of the proposed algorithm is shown in Figure 1. The audio enhancement was performed in the time–frequency domain, so short-time discrete Fourier transform (STDFT) was first performed to the audio signal. After the estimations of noise PSD and acoustic transfer function (ATF), two initial beamforming values, $\widetilde{Y}(0)$ and $\widetilde{N}(0)$, were calculated. The consistency weight matrix was used to update the beamforming value by iteration. Lastly, the enhanced audio signal could be obtained after convergence. In order to express the contribution of this paper more easily, in this section, the distributed beamformer was interpreted in detail after the centralized beamformer was introduced.

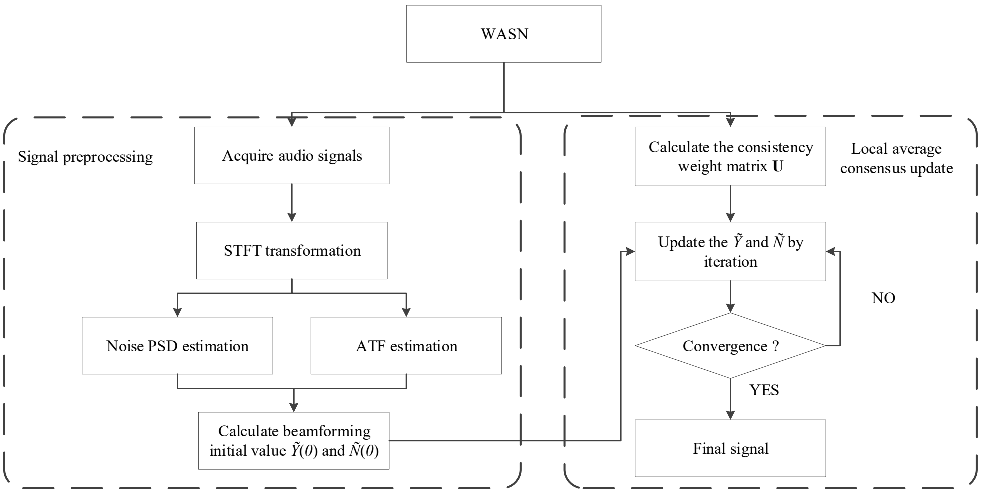

WASN
Signal preprocessing
Acquire audio signals
STFT transformation
Noise PSD estimation
ATF estimation
Calculate beamforming initial value $\tilde{Y}(0)$ and $\tilde{N}(0)$
Calculate the consistency weight matrix **U**
Local average consensus update
Update the $\tilde{Y}$ and $\tilde{N}$ by iteration
NO
Convergence ?
YES
Final signal

Figure 1. The improved distributed minimum variance distortionless response (IDMVDR) beamforming method based on the local average consensus algorithm for bird audio enhancement in wireless acoustic sensor networks (WASNs).

2.1. Centralized Beamformer (MVDR Algorithm)

In the audio enhancement of the WASN with FC, other nodes needed to broadcast their signals to the FC, so that the centralized beamformer could be obtained at FC. We considered a WASN consisting of N_d randomly connected nodes. Each node was assumed to consist of only one microphone. Here, only one signal source (bird audio) was considered. The $N_d \times 1$ vector $\mathbf{Y}(k,l)$ contains the STDFT coefficients of N_d collected signals under certain a time-frame l and frequency-bin k, which can be shown as follows:

$$\mathbf{Y}(k,l) = \mathbf{H}(k) \cdot s(k,l) + \mathbf{n}(k,l) = \mathbf{X}(k,l) + \mathbf{n}(k,l), \tag{1}$$

where the $s(k,l)$ is the STDFT coefficient of bird audio signal, and $N_d \times 1$ vector $\mathbf{n}(k,l)$ denotes the noise STDFT coefficients. The $\mathbf{H}(k)$ is a $N_d \times 1$ vector which denotes the ATFs from the bird to each microphone. If we omit the time and frequency indices for brevity, Equation (1) can be simplified as follows:

$$\mathbf{Y} = \mathbf{H} \cdot s + \mathbf{n} = \mathbf{X} + \mathbf{n}, \tag{2}$$

where vector $\mathbf{H}$ consists of N_d rows, which is given by:

$$\mathbf{H} = \left[H_1 \cdots H_{N_d} \right]^{\mathrm{T}}, \tag{3}$$

ATF is used to describe the relationship between the source signal and collected signal of microphone. ATF estimation is relatively complicated [14–16], and was not the core concern here. Considering the absence of echoes and reverberation in forest areas, the ATF in reference [9] was adopted directly in this paper, which is defined as:

$$\mathbf{H}(k) = \left[\frac{1}{\sqrt{4\Pi} q_1} e^{-2j\Pi q_1 v_f / c}, \cdots, \frac{1}{\sqrt{4\Pi} q_{N_d}} e^{-2j\Pi q_i v_f / c} \right]^{\mathrm{T}}, \tag{4}$$

when $\mathrm{j} = \sqrt{-1}$, $v_f = k \times f_s / K$ is the continuous frequency (in Hz) corresponding to the frequency bin $k \in \{0, \cdots, K/2\}$ (K denotes the STDFT length of one time frame), and f_s is the sampling frequency. q_i denotes the distance between the target bird and the node i.

The purpose of the MVDR algorithm is to keep the output signal distortionless and minimize the noise power. The MVDR algorithm is a special form of linearly constrained minimum variance (LCMV) when the source number is limited to one. With the MVDR algorithm, the STDFT coefficient of desired bird audio can be estimated by applying a complex weight to the vector **Y** with noisy STDFT coefficients. That is,

$$Z = \mathbf{W}^H \mathbf{Y} = \mathbf{W}^H \mathbf{H} s + \mathbf{W}^H \mathbf{n}, \tag{5}$$

where Z denotes the desired bird audio STDFT coefficients, **W** is a weight vector with filter coefficients, and $(\cdot)^H$ indicates Hermetian transposition of a matrix. According to the bird audio distortionless constraint, to minimize the contribution of interference to the output Z, the optimal weight vector **W** is the solution to the following optimization problem:

$$\left\{ \begin{array}{c} \mathbf{W} = arg\ min \mathbf{W}^H \mathbf{R}_{YY} \mathbf{W} \\ \mathbf{H}^H \mathbf{W} = 1 \end{array} \right. , \tag{6}$$

where $\mathbf{R}_{YY} = \mathrm{E}\left[\mathbf{Y}\mathbf{Y}^H\right]$ is the spectral covariance matrix of the noisy signal with the statistical expectation operator $\mathrm{E}[\cdot]$.

$$\mathbf{R}_{YY} = \mathrm{E}\left[\mathbf{Y}\mathbf{Y}^H\right] = \mathrm{E}\left[(\mathbf{X}+\mathbf{n})(\mathbf{X}+\mathbf{n})^H\right] = \mathrm{E}\left[\mathbf{X}\mathbf{X}^H\right] + \mathrm{E}\left[\mathbf{X}\mathbf{n}^H\right] + \mathrm{E}\left[\mathbf{n}\mathbf{X}^H\right] + \mathrm{E}\left[\mathbf{n}\mathbf{n}^H\right]. \tag{7}$$

In general, it is assumed that bird audio is uncorrelated with the noise. Therefore, the correlation coefficient is zero:

$$\mathrm{E}\left[\mathbf{X}\mathbf{n}^H\right] = \mathrm{E}\left[\mathbf{n}\mathbf{X}^H\right] = 0, \tag{8}$$

We simplified Equation (7) as follows:

$$\mathbf{R}_{YY} = \mathrm{E}\left[\mathbf{X}\mathbf{X}^H\right] + \mathrm{E}\left[\mathbf{n}\mathbf{n}^H\right] = \mathbf{R}_{XX} + \mathbf{R}_{nn}, \tag{9}$$

Because $\mathbf{H}^H\mathbf{W} = 1$,

$$\mathbf{W}^H \mathbf{R}_{XX} \mathbf{W} = \mathrm{E}\left[\mathbf{W}^H \mathbf{H} s (\mathbf{H} s)^H \mathbf{W}\right] = \mathrm{E}\left[s s^H\right], \tag{10}$$

For certain time and frequency, s is a fixed value, then the $\mathrm{E}\left[s s^H\right]$ is a fixed value too. Therefore, the optimization problem of Equation (6) can be transformed into the equations as follows:

$$\left\{ \begin{array}{c} \mathbf{W} = arg\ min \mathbf{W}^H \mathbf{R}_{nn} \mathbf{W} \\ \mathbf{H}^H \mathbf{W} = 1 \end{array} \right. , \tag{11}$$

By using the matrix inversion lemma [15], the optimal weight vector **W** of the above constrained optimization problem Equation (11) can be written as:

$$\mathbf{W} = \frac{\mathbf{R}_{nn}^{-1} \mathbf{H}}{\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}}, \tag{12}$$

Assumed that the noise coefficient of each node was spatially uncorrelated with PSD σ_i^2, then the noise correlation matrix PSD $\mathbf{R}_{nn}$ can be described by the following equation [15]:

$$\mathbf{R}_{nn} = \mathrm{E}\left[\mathbf{nn}^H\right] = diag\left\{R_{N_1N_1}, R_{N_2N_2}, \cdots, R_{N_dN_d}\right\} = diag\left\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_{N_d}^2\right\}, \tag{13}$$

$\mathbf{R}_{nn}$ will be estimated by the continuous spectrum update algorithm, which is described in Section 2.4. At last, the beamformer output is given as follows:

$$Z = \mathbf{W}^H\mathbf{Y} = \left(\frac{\mathbf{R}_{nn}^{-1}\mathbf{H}}{\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}}\right)^H \mathbf{Y} = \frac{\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} Y_i}{\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} H_i}, \tag{14}$$

where H_i and Y_i are the ATF and the collected signal of node i, respectively. This is the result of the MVDR beamforming algorithm with FC. In the centralized beamformer, the FC was required to collect the information of all sensor positions. Besides, each node had to report its Y_i to the FC. Hence, when the node number of the WASN increased, the centralized process lacked robustness and scalability.

2.2. Average Metropolis Weight Local Average Consensus Algorithm

The metropolis weight local average consensus algorithm [18] is an iterative algorithm to solve the average consensus problems in a distributed way, only the basic process of which is elaborated in this section. Its relationship with MVDR and how each node obtains the same MVDR estimated result with this method are provided in Section 2.3, with more details.

The relations between different nodes in WASNs can be represented by a topology diagram [18,19]. In order to describe the process of the consensus algorithm, some notations are needed. For the WASN with N_d nodes, we introduced an undirected graph $G = (V, \varepsilon)$, where the vertices in $V = \{1, 2, \cdots, N_d\}$ correspond to the nodes in the network, and an edge $(i, j) \in \varepsilon$ corresponds to the communication links between node i and j. The adjacency matrix $\mathrm{A} = \left[a_{ij}\right]$ indicates whether the i_{th} node and the j_{th} node are connected. If $(i, j) \in \varepsilon$, $a_{ij} = 1$, which means the nodes i and j are interconnected and communication can be executed between them, otherwise $a_{ij} = 0$. The set of neighboring nodes of node i is $C_i = \{j \in V \middle| (i, j) \in \varepsilon\}$. The degree $d_{in}(i)$ of node i is determined by the following equation:

$$d_{in}(i) = \sum_{j=1}^{N} a_{ij}, \tag{15}$$

Given the initial value $g_i(0)$ at node i (in Section 2.3, the $g_i(0)$ stands for $\tilde{\mathbf{Y}}_i(0)$ and $\tilde{\mathbf{N}}_i(0)$), the metropolis weight local average consensus algorithm aims to find the average value $g_{ave} = \frac{1}{N_d}\sum_{j=1}^{N_d} g_i(0)$ at all nodes by using an iterative scheme with only local information and local processing [14]. According to [15,16,18], the linear iterative algorithm can be used to compute the weighted mean of the error between the current value and average value, then the updated local estimate of node i is presented as follows:

$$\begin{aligned} g_i(t+1) &= g_i(t) + \sum_{j \in C_i(t)} U_{ij}(t)\left(g_j(t) - g_i(t)\right) \\ &= (1 - \sum_{j \in C_i(t)} U_{ij}(t))g_j(t) + \sum_{j \in C_i(t)} U_{ij}(t)g_i(t) \\ &= U_{ii}(t)g_i(t) + \textstyle\sum_{j \in C_i(t)} U_{ij}(t)g_j(t), \end{aligned} \tag{16}$$

where t is the iteration number, the suffix (t) means the relevant value at iteration t. $g_i(t)$ is the updated value of node i. $U_{ij}(t)$ is the updated weight between node i and j. The updated self-weight of node i, $U_{ii}(t)$ is defined as follows:

$$U_{ii}(t) = 1 - \sum_{j \in C_i(t)} U_{ij}(t), \tag{17}$$

Let the vector $\boldsymbol{g}(t) = \left[g_1(t), g_2(t), \cdots, g_{N_d}(t)\right]^{\mathrm{T}}$ denote the vector of values at iteration t. $\boldsymbol{U}(t) = \left\{U_{ij}(t)\right\}(1 \le i, j \le N_d)$ is the $N_d \times N_d$ dimensional consistency weight matrix. Equation (16) can be explained in matrix form as follows:

$$\boldsymbol{g}(t+1) = \boldsymbol{U}(t)\boldsymbol{g}(t), \tag{18}$$

The purpose of the local average consensus algorithm is to find a suitable consistency weight matrix $U(t)$, then each node will reach the average value through communicating with neighboring nodes. The average metropolis weight method can be used to fulfill the above purpose by computing the updated weight using Equation (19), which can achieve a faster convergence rate and easier implementation than the metropolis weight method [16]:

$$U_{ij}(t) = \begin{cases} \frac{2\theta}{d_{in}(i)+d_{in}(j)}\ (i,j) \in \varepsilon \\ 1 - \sum_{k \in C_i(t)} U_{ik}(t)\ \ i = j \\ 0 \text{ others} \end{cases}, \tag{19}$$

where θ $(0 < \theta < 1)$ is the trade-off factor.

2.3. Distributed MVDR (DMVDR) Beamformer

As the centralized beamformer requires an FC, there are two disadvantages of high energy consumption and poor scalability. The distributed beamformer is performed through the communications between a node and its adjacent nodes, which means that only local information and local processing are used to obtain the same optimal estimated result as Equation (14). Therefore, the distributed approach is more suitable for WASNs. In distributed beamformers, Equation (14) is rearranged as follows:

$$Z = \mathbf{W}^H\mathbf{Y} = \left(\frac{\mathbf{R}_{nn}^{-1}\mathbf{H}}{\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}}\right)^H \mathbf{Y} = \frac{\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} Y_i}{\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} H_i} = \frac{\frac{1}{N_d}\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} Y_i}{\frac{1}{N_d}\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} H_i}, \tag{20}$$

where $\frac{1}{N_d}\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} Y_i$ and $\frac{1}{N_d}\sum_{i=1}^{N_d} H_i^H \sigma_i^{-2} H_i$ can be computed by the average metropolis weight local average consensus algorithm, which means that an FC is not needed to collect the signals of all the nodes here.

Two initial values of node i are defined, namely $\widetilde{Y}_i(0) = H_i^H \sigma_i^{-2} Y_i$ and $\widetilde{N}_i(0) = H_i^H \sigma_i^{-2} H_i$. Then, Equation (20) can be written as follows:

$$Z = \frac{\frac{1}{N_d}\ \sum_{i=1}^{N_d} \widetilde{Y}_i(0)}{\frac{1}{N_d}\ \sum_{i=1}^{N_d} \widetilde{N}_i(0)}, \tag{21}$$

Let $\widetilde{Y}_{ave} = \frac{1}{N_d}\ \sum_{i=1}^{N_d} \widetilde{Y}_i(0)$ and $\widetilde{N}_{ave} = \frac{1}{N_d}\ \sum_{i=1}^{N_d} \widetilde{N}_i(0)$, Equation (20) can be written as follows:

$$Z = \frac{\widetilde{Y}_{ave}}{\widetilde{N}_{ave}}, \tag{22}$$

Then, the final purpose of the distributed MVDR (DMVDR) beamformer is to achieve the average $\widetilde{Y}_{ave}$ and $\widetilde{N}_{ave}$. Here, we used the average metropolis weight local average consensus algorithm to fulfill the distributed evaluations of $\widetilde{Y}_{ave}$ and $\widetilde{N}_{ave}$. Set $\tilde{\mathbf{Y}}(t) = \left[\widetilde{Y}_1(t), \widetilde{Y}_2(t), \cdots, \widetilde{Y}_{N_d}(t)\right]^T$ and $\tilde{\mathbf{N}}(t) = \left[\widetilde{N}_1(t), \widetilde{N}_2(t), \cdots, \widetilde{N}_{N_d}(t)\right]^T$ as the vector $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{N}}$ at iteration t, respectively. According to Equation (18), estimations of $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{N}}$ for each iteration t are given by the following equations:

$$\tilde{\mathbf{Y}}(t) = \boldsymbol{U}(t)\tilde{\mathbf{Y}}(t-1), \tag{23}$$

$$\tilde{\mathbf{N}}(t) = \boldsymbol{U}(t)\tilde{\mathbf{N}}(t-1), \tag{24}$$

Let $\widetilde{Z}_i(t)$ denote the distributed MVDR output at iteration t, which can be calculated by using the following equation:

$$\widetilde{Z}_i(t) = \widetilde{Y}_i(t)/\widetilde{N}_i(t), \tag{25}$$

2.4. Noise PSD Estimation Algorithm Based on Continuous Spectrum Update

In order to calculate the initial values $\widetilde{Y}_i(0)$ and $\widetilde{N}_i(0)$, the noise PSD $\sigma_i^2(k,l)$ is required. An accurate estimation of the noise PSD substantially affects the quality of the final output signal in distributed MVDR beamformers. The matrix of noise PSD is typically estimated by the temporal averaging of part of the audio with no bird vocalization [10,20,21]. The premise is that the noise PSD does not change too much with time [21]. However, noises usually vary with time in nature, which makes it a great challenge to accurately estimate the noise PSD. Compared with the noise estimation algorithm based on voice activity Ddetection (VAD), the algorithm based on continuous updates is more effective. These algorithms are usually based on the minimum statistical model, which consider that the point with the smallest amplitude is the reference value of noise estimation for continuous multiple frames with the same spectral component [22,23]. Based on this model, the noise PSD can be estimated. However, if only the current frame is searched, this will cause large residual noise when the noise level rises soon afterwards [17]. In this section, an improved noise PSD estimation algorithm is proposed to estimate the noise PSD in each node during IDMVDR. This method learned from the bidirectional search of the path searching algorithm.

Firstly, the smoothing operation on the power spectrum of the signal of node i is performed:

$$S_i(k,l) = \alpha' S_i(k,l-1) + (1-\alpha')\left|Y_i(k,l)\right|^2, \tag{26}$$

where $S_i(k,l)$ denotes the smoothed power spectrum, and α' is the smoothing factor.

Then, the bidirectional search is performed on the smoothed power spectrum. In the same spectral component, the front and rear frames are searched to find the smallest amplitude:

$$S_{imin1}(k,l) = \min\{S_i(k,l')\}, l-L_s+1 \le l' \le l, \tag{27}$$

$$S_{imin2}(k,l) = \min\{S_i(k,l')\}, l \le l' \le l+L_s-1, \tag{28}$$

where $S_{imin1}(k,l)$ and $S_{imin2}(k,l)$ represent the values of forward and backward searches of node i, respectively, and L_s is the number of searched frames.

The maximums of S_{imin1} and S_{imin2} are taken as the reference values, which can ensure the reduction of noise when the noise level rises by utilizing the noise reference values of later frames:

$$S_{imin}(k,l) = \max\{S_{imin1}(k,l), S_{imin2}(k,l)\}, \tag{29}$$

Furthermore, the bird vocalization existence probability is calculated. Considering that the bird vocalization is more susceptible to noise interference in the low-frequency segment, different discrimination thresholds are chosen for signals with different frequency bands:

$$I_i(k,l) = \begin{cases} 1, S_i(k,l)/S_{imin}(k,l) > \delta(k) \\ 0, S_i(k,l)/S_{imin}(k,l) \le \delta(k) \end{cases}, \tag{30}$$

where $I_i(k,l)$ is used to determine whether bird vocalization is present in time-frame l and frequency-bin k. When the value of $I_i(k,l)$ is 1, it means that bird vocalization is present, otherwise bird vocalization is absent. $\delta(k)$ is the frequency-dependent threshold determined experimentally. The existence probability of bird vocalization is further calculated by the following equation:

$$P_i(k,l) = \alpha_s P_i(k,l-1) + (1-\alpha_s) I_i(k,l), \tag{31}$$

where $P_i(k,l)$ is the existence probability of bird vocalization, and α_s is smoothing constant.

To quickly adapt to the increasing noise levels, the time–frequency dependent smoothing factor is computed based on the existence probability of bird vocalization. Then, the noise estimation will be updated in each frame based on the activity detection of bird vocalization. Let $\alpha'_{i,d}(k,l)$ denote the time–frequency dependent smoothing factor of node i, it can be calculated by the following equation:

$$\alpha'_{i,d}(k,l) = \alpha_d + (1-\alpha_d) P_i(k,l), \tag{32}$$

where α_d is the fixed smoothing factor of $\alpha'_{i,d}(k,l)$.

Let $S_f(k,l)$ describe the variation of noise power spectrum, the $S_f(k,l)$ of node i is computed by the following equation:

$$S_{i,f}(k,l) = S_{imin1}(k,l) / S_{imin2}(k,l), \tag{33}$$

When the noise power spectrum varies little, $S_{imin}(k,l)$ is used to estimate noise PSD, otherwise the noise PSD is estimated with the time–frequency dependent smoothing factor $\alpha'_{i,d}(k,l)$. Then, the noise PSD is estimated through by the following equation:

$$\sigma_i^2(k,l) = \begin{cases} S_{imin}(k,l), \ 1/c < S_{i,f}(k,l) < c \\ \alpha'_{i,d}(k,l)\sigma_i^2(k,l-1) + \left[1-\alpha'_{i,d}(k,l)\right]\left|Y(k,l)\right|^2, \text{ other} \end{cases}, \tag{34}$$

where the $\sigma_i^2(k,l)$ is the noise PSD of node i, and c is the decision threshold constant.

2.5. The Iterative Procedure of IDMVDR

In summary, the iterative procedures of IDMVDR are listed as follows:

(1) Initialize iteration t to 0, calculate $\widetilde{\mathbf{Y}}_i(0)$ and $\widetilde{\mathbf{N}}_i(0)$ by using $\widetilde{Y}_i(0) = H_i^H \sigma_i^{-2} Y_i$ and $\widetilde{N}_i(0) = H_i^H \sigma_i^{-2} H_i$, where H_i is obtained by Equation (4), and σ_i^2 is computed by Equation (34);

(2) Use Equation (15) to compute the degree of each node based on the current topology diagram, then compute the current consistency weight matrix $U(t)$ by using Equation (19);

(3) In each node, the updated weights $U(t-1)$, $\widetilde{Y}(t-1)$, and $\widetilde{N}(t-1)$ of itself and its adjacent nodes are applied to calculate $\widetilde{Y}(t)$, $\widetilde{N}(t)$ and $\widetilde{Z}(t)$ by using Equations (23)–(25), then increase t by 1;

(4) Repeat step (3), until the convergence of $\widetilde{Z}_i(t)$.

Significantly, the detailed communication contents between adjacent nodes are only the updated weights $U(t-1)$, $\widetilde{Y}(t-1)$ and $\widetilde{N}(t-1)$.

Take a WASN with three nodes as an example, the topology is shown in Figure 2. Here, yellow dots are the nodes and blue lines denote the communication connections between nodes.

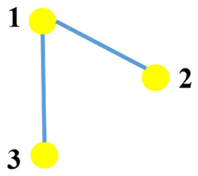


Figure 2. The topology of a WASN with three nodes.

Assuming that the θ of Metropolis is 0.5, then the weight between node 1 and 2 is 1/(2 + 1) = 1/3, according to Equation (19), and the weight between node 1 and 3 is 1/(2 + 1) = 1/3. The self-weight of node 1 is 1 − 1/3 − 1/3 = 1/3 by using Equation (17). Node 2 and node 3 are not connected, so the weight between node 2 and 3 is 0. The self-weights of node 2 and node 3 are both 1 − 1/3 = 2/3. The consistency weight matrix $U(t)$ of this WASN is as follows:

$$U(\mathrm{t}) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \end{bmatrix}, \tag{35}$$

$\tilde{\mathbf{Y}}(t) = \left[\widetilde{Y}_1(t), \widetilde{Y}_2(t), \widetilde{Y}_3(t)\right]^T$ and $\tilde{\mathbf{N}}(t) = \left[\widetilde{N}_1(t), \widetilde{N}_2(t), \widetilde{N}_3(t)\right]^T$ are the vectors of all nodes' values at iteration t respectively. So, for the iteration t + 1,

$$\begin{aligned} \tilde{\mathbf{Y}}(t+1) = \boldsymbol{U}(\mathrm{t})\tilde{\mathbf{Y}}(t) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \end{bmatrix} \left[\widetilde{Y}_1(t), \widetilde{Y}_2(t), \widetilde{Y}_3(t)\right]^T \\ = \left[\tfrac{1}{3}\widetilde{Y}_1(t) + \tfrac{1}{3}\widetilde{Y}_2(t) + \tfrac{1}{3}\widetilde{Y}_3(t), \tfrac{1}{3}\widetilde{Y}_1(t) + \tfrac{2}{3}\widetilde{Y}_2(t), \tfrac{1}{3}\widetilde{Y}_1(t) + \tfrac{2}{3}\widetilde{Y}_3(t)\right]^T, \end{aligned} \tag{36}$$

with $\tilde{\mathbf{N}}(t)$, $\tilde{\mathbf{N}}(t+1)$ can be obtained by the same method. As for node 1, its $\widetilde{Z}_1(t)$ can be computed as follows by using Equation (25):

$$\widetilde{Z}_1(t+1) = \frac{\widetilde{Y}_1(t+1)}{\widetilde{N}_1(t+1)} = \frac{\frac{1}{3}\widetilde{Y}_1(t) + \frac{1}{3}\widetilde{Y}_2(t) + \frac{1}{3}\widetilde{Y}_3(t)}{\frac{1}{3}\widetilde{N}_1(t) + \frac{1}{3}\widetilde{N}_2(t) + \frac{1}{3}\widetilde{N}_3(t)}, \tag{37}$$

The iteration would be executed until the difference between $\widetilde{Z}(t+1)$ and $\widetilde{Z}(t)$ is less than a given threshold.

3. Comparison and Discussion

In this section, we provide qualitative comparisons between our proposed method and two other existing audio enhancement methods (DDSB and DMVDR). Both convergence rate and noise reduction performance were used to conclude the advantages of our proposed method.

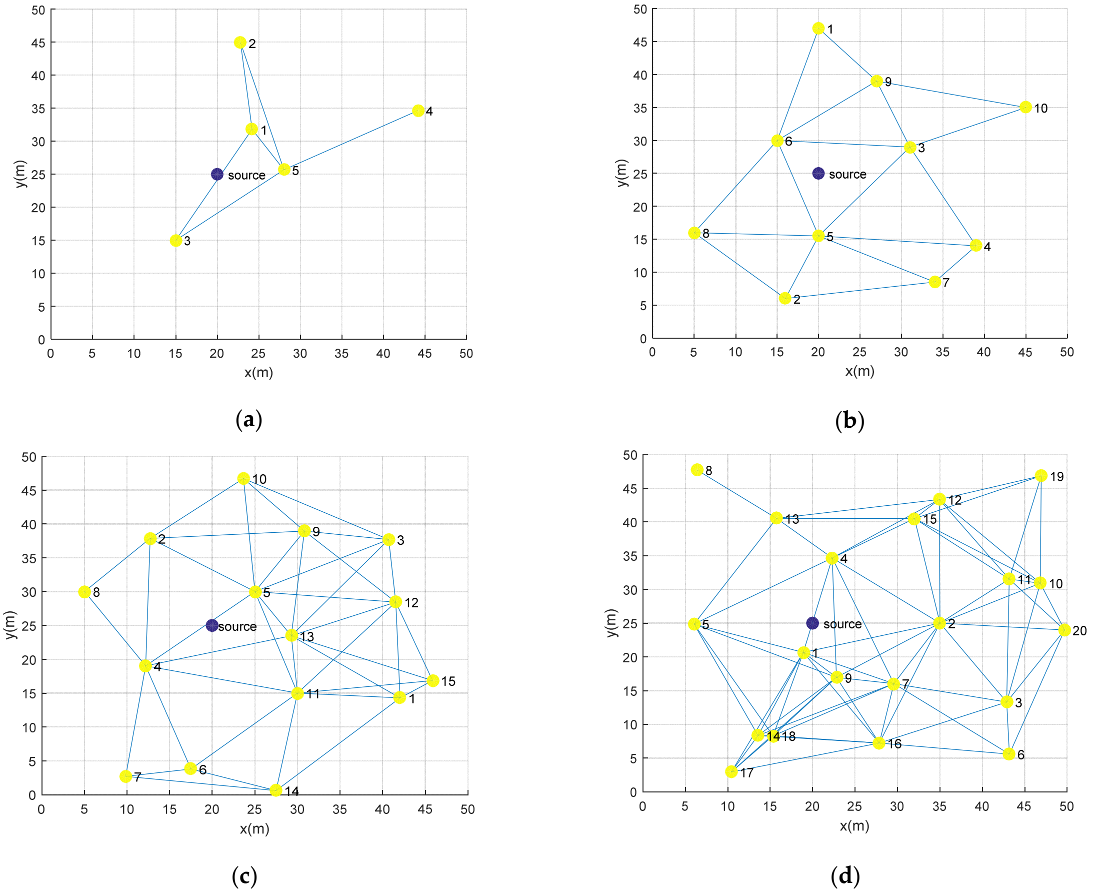
3.1. Convergence Analysis

Different network topologies of WASNs will cause different convergence and enhancement performances. To evaluate the performance of WASNs of different sizes, four WASNs with N_d = 5, 10, 15, and 20 were simulated. All the nodes were distributed in a 50 × 50 m site and the heights were 1 m. Each node consisted of one microphone. The topologies of four simulated WASNs were randomly generated. A bird vocalization source was fixed at a point 5 m high. The noise source was randomly distributed. The topologies of the four simulated WASNs are shown in Figure 3a–d, respectively. In these figures, yellow dots are the node position of WASNs, dark blue dots are the position of the bird vocalization source, blue lines denote the communication connections between nodes.

In the simulation experiments, bird audio of a large-billed crow with relatively high signal-to-noise ratio (SNR) was selected as the clean bird vocalization. Its waveform is shown in Figure 4.

In this section, we only verify the convergence rate of IDMVDR. Before the iterative update of the node value, the estimation of the noise PSD was completed, so the convergence rate of the distributed algorithm had nothing to do with the noise type. A white Gaussian signal was chosen as the only noise source. Each node collected noisy bird audio signals with a sampling frequency of 44.1 kHz. To guarantee the same conditions for comparison experiments, the input SNR of node 1 was set to 1 dB.

The audio signal was non-stationary, which should be framed and windowed before audio enhancement. All nodes processed the signals frame by frame, with a frame length of 706 (16 ms), 50% overlap, and Hanning window. During the noise PSD estimation, the first and second frame signals were first initialized through setting $\sigma_i^2(k,l) = S_i(k,l) = S_{imin}(k,l) = \left|Y_i(k,l)\right|$, $P_i(k,l) = 0$. Then, the noise PSD estimation was performed, beginning with the second frame, by equations from Equation (26) to Equation (34) sequentially, where $\alpha' = 0.8$ (in Equation (26)), $L_s = 95$ (in Equations (28) and (29)), $\alpha_s = 0.2$ (in Equation (31)), $\alpha_d = 0.95$ (in Equation (32)) and $c = 4$ (in Equation (34)).



(**a**)

(**b**)

(**c**)

(**d**)

Figure 3. The topologies of four WASNs with different node numbers: (**a**) 5 nodes; (**b**) 10 nodes; (**c**) 15 nodes; (**d**) 20 nodes.

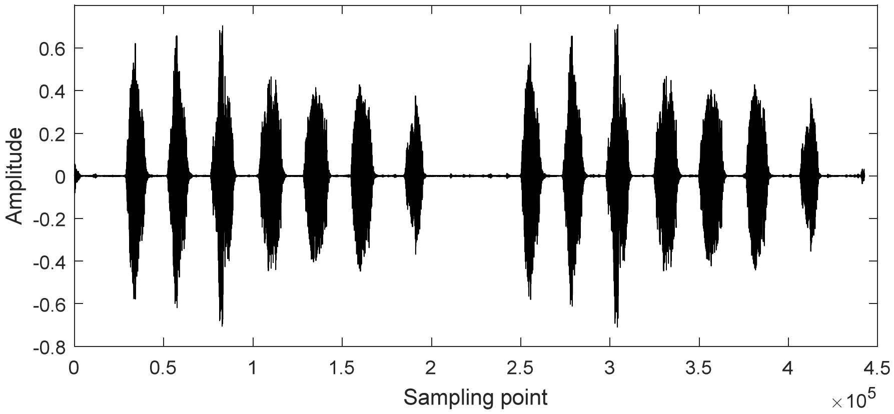


Figure 4. Signal of clean bird vocalization.

Considering the characteristics of different kinds of bird vocalization are different, relatively appropriate decision thresholds for the bird vocalizations were selected. After statistical analysis, we found that most of the bird vocalizations were in medium and low frequencies. Through lots of experiments with different kinds of birds, the decision threshold of the medium and low frequency segment was set to 2, which makes it easier to confirm that the bird vocalization exists and reduce the distortion of bird vocalization in medium and low frequency. The upper limit of the medium frequency was selected as 16.61 kHz, and the corresponding frequency-bin k was 266 and 440. For the high frequency segment, the decision threshold was set to 5. Thus, the value of $\delta(k)$ could be set as follows:

$$\delta(k) = \begin{cases} 5, \ 266 < k < 440 \\ \quad 2, \ \text{other} \end{cases}, \tag{38}$$

The performance of convergence is the main concern in this section, and only the MSE is introduced herein as the measurement of differences between source signal and estimated signal, which is shown as follows [16]:

$$\text{MSE} = \frac{1}{LK} \sum_{l=1}^{L} \sum_{k=1}^{K} \left| \hat{Z}_i(k,l) - S(k,l) \right|^2, \tag{39}$$

where K denotes the number of frequency bins, L is the number of time-frames, $\hat{Z}_i(k,l)$ and $S(k,l)$ denote the STDFT coefficients of the beamformer output and the desired signal at frequency-bin index k and time-frame index l, respectively. During the iteration process of IDMVDR, the MSE changes. When the MSE does not change any more, it represents the convergence of IDMVDR. The criteria for iteration stop are shown as follows:

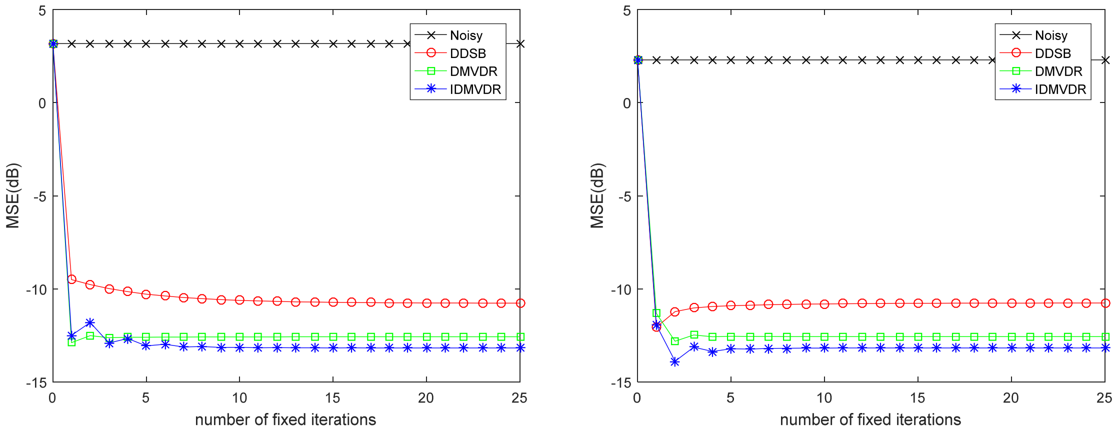
$$\text{MSE}(t+1) - \text{MSE}(t) = 10\text{E}-6, \tag{40}$$

where MSE(t) is the value of MSE at iteration t.

Convergence rate is generally measured by the number of iterations when converging is used [20]. To verify the faster convergence rate of IDMVDR, two other methods, DDSB and DMVDR [9,14,15,20], were compared in four different WASNs.

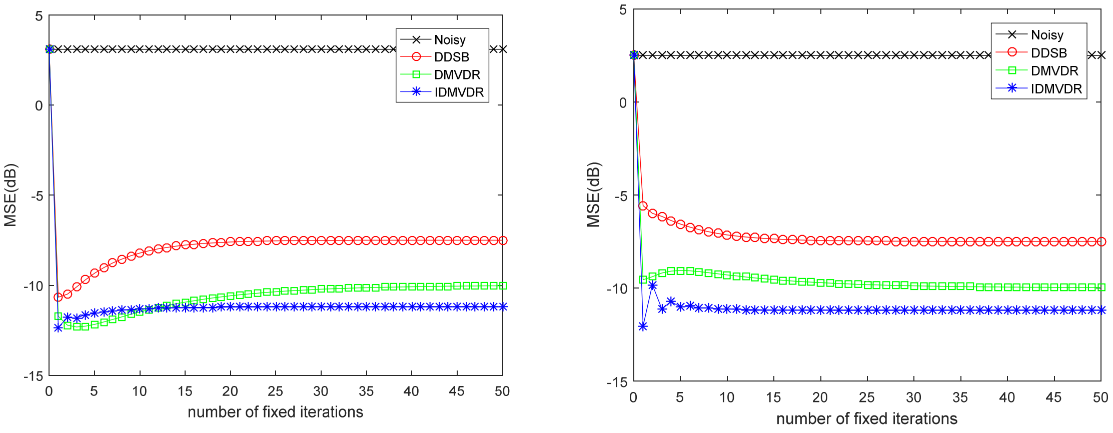
For each WASN, only the results of two nodes with maximum and minimum degrees were analyzed. The MSE variations in WASNs with N_d = 5, 10, 15, 20 are shown in Figures 5–8, respectively, the corresponding numbers of iterations when converging are shown in Tables 1–4.

When N_d = 5, 10, and 15, as shown in Figures 5–7 and Tables 1–3, the convergence rate of IDMVDR was relatively faster than that of DMVDR and DDSB. Meanwhile, IDMVDR could converge to a lower MSE. Moreover, in some cases, the convergence rate of DDSB was faster than DMVDR, but IDMVDR was always the fastest one.



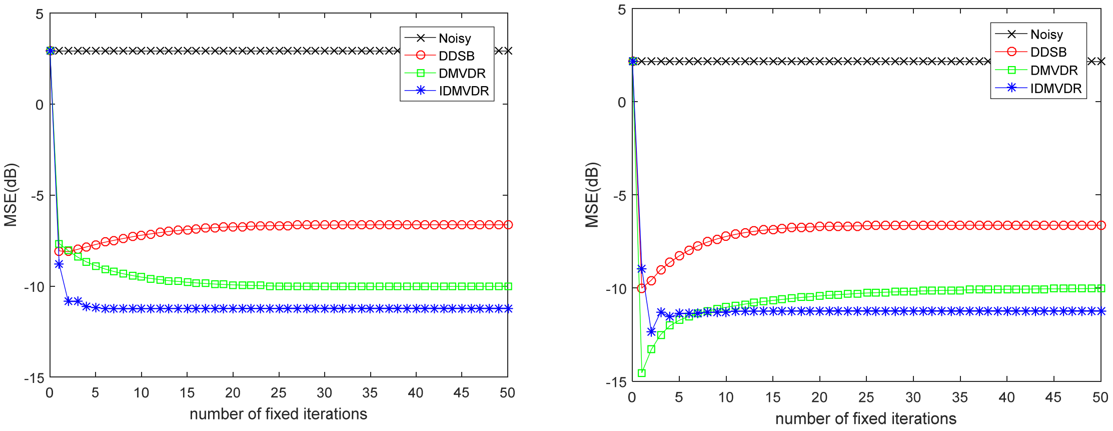
(**a**) (**b**)

Figure 5. The MSE variations in the WASN with N_d = 5: (**a**) MSE of node 4 in each iteration; (**b**) MSE of node 5 in each iteration.



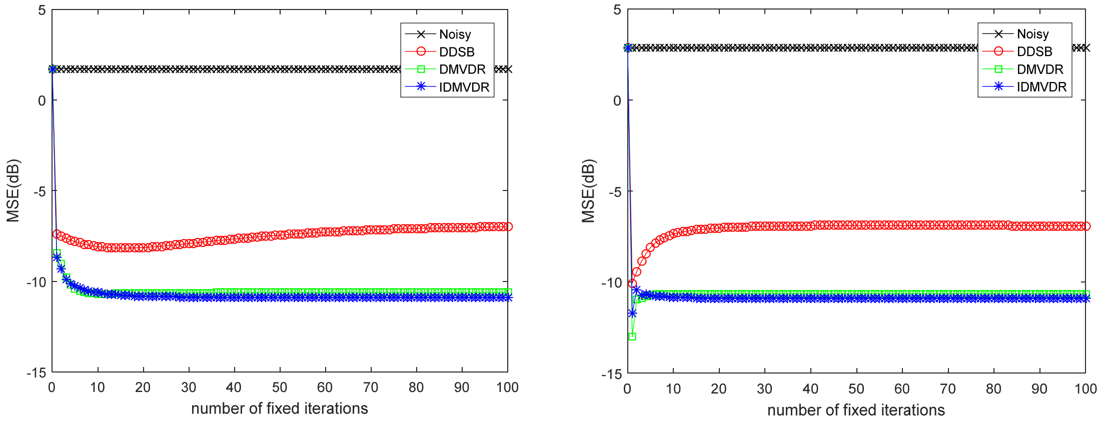
(**a**) (**b**)

Figure 6. The MSE variations in the WASN with N_d = 10: (**a**) MSE of node 1 in each iteration; (**b**) MSE of node 5 in each iteration.



(**a**) (**b**)

Figure 7. The MSE variations in the WASN with N_d = 15: (**a**) MSE of node 8 in each iteration; (**b**) MSE of node 5 in each iteration.



(**a**) (**b**)

Figure 8. The MSE variations in the WASN with N_d = 20: (**a**) MSE of node 8 in each iteration; (**b**) MSE of node 2 in each iteration.

Table 1. The numbers of iterations when converging in the WASN with $N_d = 5$.

Algorithm	Node 4	Node 5
DDSB	15	5
DMVDR	11	4
IDMVDR	6	2

Table 2. The number of iterations when converging in the WASN with $N_d = 10$.

Algorithm	Node 1	Node 5
DDSB	35	33
DMVDR	38	30
IDMVDR	19	11

Table 3. The number of iterations when converging in the WASN with $N_d = 15$.

Algorithm	Node 8	Node 5
DDSB	37	28
DMVDR	49	43
IDMVDR	10	6

Table 4. The number of iterations when converging in the WASN with $N_d = 20$.

Algorithm	Node 8	Node 2
DDSB	95	45
DMVDR	64	22
IDMVDR	51	19

When $N_d = 20$, as shown in Figure 8 and Table 4, the convergence rate of IDMVDR was also significantly faster than that of DDSB, and the MSE of IDMVDR when converging was far less than DDSB. The MSE variation curves of IDMVDR and DMVDR almost completely overlap, which means that the performances were nearly the same. This is different from the above results of $N_d = 5$, 10, and 15. This might be attributable to the higher node density, which means more nodes can get high SNR signals, and then the difference between IDMVDR and DMVDR is not obvious.

As for convergence performance, compared with DMVDR and DDSB, IDMVDR was predominant in all the four WASNs, and the convergence rate of IDMVDR was always the fastest, which means that less data transmissions were required in IDMVDR, and less energy was consumed to communicate further with neighboring nodes. Thus, the IDMVDR method can save more energy than DDSB and DMVDR, which is more suitable for the WASNS with limited energy supply.

3.2. Noise Reduction Performance Analysis

Three different indexes, MSE, SegSNR, and PESQ, were selected as the measurements of the noise reduction performance. The SegSNR is averaged over all time frames and is given by [15]:

$$\text{SegSNR} = \frac{1}{L}\sum\nolimits_{l=1}^{\mathrm{L}} 10log_{10}\frac{\sum_{k=1}^{\mathrm{K}}\left|S(k,l)\right|^2}{\sum_{k=1}^{\mathrm{K}}\left|\hat{Z}_i(k,l)-S(k,l)\right|^2}, \tag{41}$$

PESQ is the most common and subjective metric for evaluating speech quality, which can be used in a wide range of end to end measurement applications with live and simulated networks [24]. PESQ is a more complex metric for capturing a wider range of distortions. Its value varies from −0.5 to 4.5, and a higher value means the better performance. Here, PESQ was introduced to measure the quality of enhanced bird vocalization. The detailed calculation method can be seen in [24].

Also, the four WASNs in Section 3.1 were applied for simulation. One bird vocalization and ten types of noises were selected as audio sources for each experiment. Here, bird vocalizations were high SNR, and selected from Freesound [25], xeno-canto [26], or our laboratory. The ten types of noises contained five independent white Gaussian noise signals and five independent other noise signals, including three types of noise from the Aurora2 database [27], namely Babble, Restaurant, and Street. In all the experiments, the input SNR of node 1 was set to 1 dB. Ten experiments with different audio sources were performed for all three methods, and the final results are the average of multiple experiments.

Tables 5 and 6 show the MSEs and SegSNRs of the distributed beamformers, respectively. It was found that the noise reduction performance of IDMVDR was better than that of DMVDR and DDSB, which means that the proposed PSD estimation algorithm of IDMVDR is more efficient.

Table 5. The noise reduction performance of different N_d in terms of MSE.

WASN of Different N_d	DDSB	DMVDR	IDMVDR
5	−10.74 dB	−12.56 dB	−13.17 dB
10	−7.49 dB	−9.99 dB	−11.13 dB
15	−6.63 dB	−10.00 dB	−11.21 dB
20	−6.89 dB	−10.63 dB	−10.87 dB
Ave	−7.94 dB	−10.79 dB	−11.85 dB

Table 6. The noise reduction performance of different N_d in terms of SegSNR.

WASN of Different N_d	DDSB	DMVDR	IDMVDR
5	−5.56 dB	−5.44 dB	−3.05 dB
10	−6.49 dB	−6.45 dB	−3.31 dB
15	−7.12 dB	−6.67 dB	−3.24 dB
20	−6.98 dB	−6.32 dB	−4.67 dB
Ave	−6.54 dB	−6.22 dB	−3.68 dB

Table 7 shows the PESQs of the beamformer's output. It was observed that the PESQs of the IDMVDR and DMVDR are obviously better than that of DDSB. This is reasonable, since the MVDR beamformer generally had better speech quality and intelligibility than the DDSB algorithm when the noise signals of the microphones were correlated [15]. Meanwhile, the bird vocalization quality of IDMVDR was slightly better than that of DMVDR.

Table 7. The noise reduction performance of different N_d in terms of PESQ.

WASN of Different N_d	DDSB	DMVDR	IDMVDR
5	1.89	2.47	2.67
10	1.55	2.11	2.50
15	1.56	2.14	2.49
20	1.58	2.18	2.48
Ave	1.65	2.22	2.54

4. Conclusions

In this paper, taking into consideration the highly time-variable noises in nature and the limited energy, calculation ability, and communication ability of wireless sensor nodes, the IDMVDR beamforming method for bird audio enhancement in WASNs was proposed. The continuous spectrum update algorithm was used to estimate the noise PSD, and the average metropolis weight consensus algorithm was introduced to fasten the convergence rate of each node herein. To validate the advantages of the proposed method, the audio enhancement experiments in four WASNs with different network topologies were simulated in MATLAB. Compared with DDSB and DMVDR, the average MSE, SegSNR, and PESQ of the enhanced signals obtained by the proposed method were best. Additionally, the proposed algorithm presented a faster convergence rate, which provides the capability of reducing the iteration time and communication cost of the network. Thanks to the above better performances, the proposed IDMVDR method is more suitable for practical utilization in WASNs with variable topology. Through the proposed method, better quality bird audio files can be achieved in bird audio monitoring, which is more advantageous for improving the accuracy of bird species identification. However, there was a simplifying assumption in our current research: only one bird audio source was considered. In nature, there is always more than one bird singing or vocalizing at the same time. In this complex situation, we need to do further research on audio enhancement.

Author Contributions: J.Z. and J.X. proposed the algorithm and designed the experiments; W.B. and Z.X. provided and processed the sound data; X.L. and B.Z. performed the experiments; B.Z. and J.X. analyzed the data; all the authors participated in the paper writing.

Funding: The paper is supported by "Science and Technology Project of State Grid Corporation of China (SGGR0000WLJS1801082)" and "the Fundamental Research Funds for the Central Universities (No. 2016ZCQ08 & No. 2015ZCQ-GX-04)".

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

References

1. Stattner, E.; Collard, M.; Hunel, P. Acoustic Scheme to Count Bird Songs with Wireless Sensor Networks. In Proceedings of the 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Lucca, Italy, 20–24 June 2011; pp. 1–3.
2. Boulmaiz, A.; Messadeg, D.; Doghmane, N.; Taleb-Ahmed, A. Robust acoustic bird recognition for habitat monitoring with wireless sensor networks. *Int. J. Speech Technol.* **2016**, *19*, 631–645. [CrossRef]
3. Xiaomin, Z.; Ying, L.I. Bird sounds recognition based on Radon and translation invariant discrete wavelet transform. *J. Comput. Appl.* **2014**, *34*, 1391–1396.
4. Attabi, Y.; Chung, H.; Champagne, B.; Zhu, W.P. NMF-based speech enhancement using multitaper spectrum estimation. In Proceedings of the International Conference on Signals and Systems, Bali, Indonesia, 1–3 May 2018; pp. 36–41.
5. Loizou, P.C. *Speech Enhancement: Theory and Practice*; Engineering & Technology: Boca, Raton, 2007; p. 632.
6. Alías, F.; Socoró, J.C.; Sevillano, X. A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Appl. Sci.* **2016**, *6*, 143. [CrossRef]
7. Lu, C.T.; Lei, C.L.; Shen, J.H.; Wang, L.L.; Tseng, K.F. Estimation of Noise Magnitude for Speech Denoising Using Minima-Controlled-Recursive-Averaging Algorithm Adapted by Harmonic Properties. *Appl. Sci.* **2017**, *7*, 9. [CrossRef]
8. Markovich, S.; Gannot, S.; Cohen, I. Multichannel Eigenspace Beamforming in a Reverberant Noisy Environment with Multiple Interfering Speech Signals. *IEEE Trans. Audio Speech Lang. Process.* **2009**, *17*, 1071–1086. [CrossRef]
9. Gannot, S.; Vincent, E.; Markovich-Golan, S.; Ozerov, A. A Consolidated Perspective on Multimicrophone Speech Enhancement and Source Separation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 692–730. [CrossRef]
10. Markovich-Golan, S.; Bertrand, A.; Moonen, M.; Gannot, S. Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks. *Signal. Process.* **2015**, *107*, 4–20. [CrossRef]
11. Hassani, A.; Bertrand, A.; Moonen, M. Cooperative integrated noise reduction and node-specific direction-of-arrival estimation in a fully connected wireless acoustic sensor network. *Signal. Process.* **2015**, *107*, 68–81. [CrossRef]
12. Bertrand, A.; Moonen, M. Distributed Adaptive Estimation of Node-Specific Signals in Wireless Sensor Networks with a Tree Topology. *IEEE Trans. Signal. Process.* **2011**, *59*, 2196–2210. [CrossRef]
13. Hassani, A.; Plata-Chaves, J.; Bahari, M.H.; Moonen, M.; Bertrand, A. Multi-Task Wireless Sensor Network for Joint Distributed Node-Specific Signal Enhancement, LCMV Beamforming and DOA Estimation. *IEEE J. Sel. Top. Signal. Process.* **2017**, *11*, 518–533. [CrossRef]

14. Zeng, Y.; Hendriks, R.C. Distributed delay and sum beamformer for speech enhancement in wireless sensor networks via randomized gossip. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal, Kyoto, Japan, 25–30 March 2012; pp. 4037–4040.
15. Zeng, Y.; Hendriks, R.C. Distributed Delay and Sum Beamformer for Speech Enhancement via Randomized Gossip. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 260–273. [CrossRef]
16. Li, D. Research on Distributed Speech Enhancement Methods in Wireless Acoustic Sensor Networks. Ph.D. Thesis, Dalian University of Technology, Dalian, China, 2015.
17. Liu, X.; Gao, Y. Speech Enhancement Algorithm with Leading-in Delay. *Mod. Electron. Technol.* **2011**, *34*, 85–88.
18. Avrachenkov, K.; Chamie, M.E.; Neglia, G. A local average consensus algorithm for wireless sensor networks. In Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems and Workshops, Barcelona, Spain, 27–29 June 2011; pp. 1–6.
19. Tian, F.F. Research on Consistent Filtering Algorithm for Wireless Sensor Networks. Ph.D. Thesis, Jiangnan University, Jiangnan, China, 2015.
20. Kodrasi, I.; Doclo, S. Joint Late Reverberation and Noise Power Spectral Density Estimation in a Spatially Homogeneous Noise Field. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal, Calgary, AB, Canada, 15–20 April 2018; pp. 441–445.
21. Liang, Y.U.; Wu, H.J.; Jiang, W.K. Multi-channel Speech Enhancement based on Beamforming and GAN Network. In *Noise & Vibration Control*; Wiley: Hoboken, NJ, USA, 2018.
22. Rangachari, S.; Loizou, P.C. A noise-estimation algorithm for highly non-stationary environments. *Speech Commun.* **2006**, *48*, 220–231. [CrossRef]
23. Fahim, A.; Samarasinghe, P.N.; Abhayapala, T.D. PSD Estimation and Source Separation in a Noisy Reverberant Environment Using a Spherical Microphone Array. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1594–1607. [CrossRef]
24. Rix, A.; Beerends, J.; Hollier, M.; Hekstra, A. *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*; ITU-T Recommendation: Geneva, Switzerland, 2001; p. 862.
25. Repository of Sound Under the Creative Commons License. Available online: http://www.freesound.org/ (accessed on 18 November 2018).
26. Sarasa, G.; Granados, A.; Rodriguez, F.B. An Approach of Algorithmic Clustering Based on String Compression to Identify Bird Songs Species in Xeno-Canto Database. In Proceedings of the International Conference on Frontiers of Signal, Paris, France, 6–8 September 2017; pp. 101–104.
27. Hirsch, H.G.; Pearce, D. The AURORA Experimental Framework for the Preformance Evaluations of Speech Recognition Systems Under Noisy Conditions. In Proceedings of the ISCA ITRW ASR, Paris, France, 18–20 September 2000; pp. 181–188.