

DOI: 10.13245/j.hust.250322

# 面向单传声器节点的分布式语音增强算法

庞淑蓉<sup>1</sup> 秦品乐<sup>1</sup> 曾建潮<sup>1</sup> 畅瑞江<sup>2</sup>

(1. 中北大学计算机科学与技术学院, 山西 太原 030051;

2. 河北建筑工程学院电气工程学院, 河北 张家口 075000)

**摘要** 在节点为单传声器的无线声学传感器网络(WASN)中, 针对分布式语音增强时声学传递函数难以获取及节点间频繁交换数据的问题, 提出了一种基于单传声器节点的分布式语音增强算法. 具体地, 首先针对全连接WASN, 提出了基于最大输入信噪比节点的语音失真加权分布式维纳滤波(SDW-DWF)算法. 该算法通过计算获得最大输入信噪比节点, 并由该节点接收其他节点发送的多帧信号信息, 来分布式实现加权维纳滤波语音增强. 该算法在节点间仅需交换少量数据, 且不依赖于声学传递函数. 此外, 提出了一种无拓扑约束的比较一致性算法, 来获得最大输入信噪比节点, 以使得SDW-DWF方法可用于无网络拓扑约束的移动说话人场景. 仿真和实际实验验证了本文方法在无需先验约束情况下, 具有较高的分布式语音增强性能.

**关键词** 分布式语音增强; 分布式维纳滤波; 加权维纳滤波; 一致性算法; 无线声学传感器网络

**中图分类号** TN912.35 **文献标志码** A **文章编号** 1671-4512(2025)03-0099-10

## Distributed speech enhancement with single microphone per node

PANG Shurong<sup>1</sup> QIN Pinle<sup>1</sup> ZENG Jianchao<sup>1</sup> CHANG Ruijiang<sup>2</sup>

(1. School of Computer Science and Technology, North University of China, Taiyuan 030051, China;

2. School of Electrical Engineering, Hebei University of Architecture, Zhangjiakou 075000, Hebei China)

**Abstract** To address the issues of difficult acoustic transfer function acquisition and frequent data exchange among nodes for distributed speech enhancement, a distributed speech enhancement method was proposed for the wireless acoustic sensor network (WASN) with a single microphone per node. Specifically, based on the node with the maximum input signal-to-noise ratio, the speech distortion weighted distributed Wiener filter (SDW-DWF) was first proposed in the fully connected WASN. To achieve distributed weighted Wiener filtering speech enhancement, the maximum input signal-to-noise ratio node needs to be obtained, and the node receives multiple frames of signal information sent by other nodes. The proposed method only requires data exchange among nodes once and does not rely on the acoustic transfer function. In addition, to obtain the node with the maximum input signal-to-noise ratio in any topology, the comparative consensus algorithm was proposed, so that the SDW-DWF method can be used in mobile speaker scenarios without topology constraint. Simulation and real-world experimental results illustrate that the proposed method without any prior information achieves good speech enhancement performance.

**Key words** distributed speech enhancement; distributed Wiener filter; weighted Wiener filter; consensus algorithm; wireless acoustic sensor network

广泛应用的语音相关设备, 例如电话会议、助听器, 收集到的音频信号性能往往会受到背景噪声的影响. 为提高音频信号的质量, 学者们先后提

出了大量基于单传声器或传声器阵列的语音增强方法<sup>[1-4]</sup>. 近些年, 随着物联网的发展, 越来越多的电子设备可以实现无线通信, 让一些配有传声器的设

**收稿日期** 2024-05-15.

**作者简介** 庞淑蓉(1987—), 女, 讲师; 畅瑞江(通信作者), 讲师, E-mail: crj2293@hebiace.edu.cn.

**基金项目** 山西科技厅重大专项基金资助项目(202101010101018); 山西省基础研究计划青年项目(202403021212105); 河北建筑工程学院博士启动基金资助项目(B-202406).

备进行连接,即可组成无线声学传感器网络(wireless acoustic sensor network, WASN). 因为 WASN 的覆盖范围大,所以节点间的间距通常大于语音信号的波长,导致无法满足空间采样定理,使得传统语音增强技术无法直接应用于 WASN. 为此,须要根据 WASN 的特点研究适用于其自身的语音增强技术.

近年来,许多学者为了在 WASN 中实现分布式语音增强,提出很多与分布式语音增强相关的算法. 他们针对不同的网络拓扑结构,提出了对应的分布式求解方法,其中包括全连接网络拓扑、树型网络拓扑、任意网络拓扑等<sup>[5-12]</sup>. 具体地,文献[5]基于线性最小均方误差(linear minimum mean squared error, LMMSE)准则,在全连接网络中首次提出了分布式自适应节点特定信号估计(distributed adaptive node-specific signal estimation, DANSE)方法. 该方法以节点顺序循环更新的方式,完成分布式信号估计任务,且更新节点仅利用其他节点的压缩信号,就可以使输出信号收敛至集中式解. 同时,文献[6]提出了基于节点同步更新的 DANSE 方法,进一步加快了估计信号收敛至集中式解的速度. 随后,文献[7]将 DANSE 方法扩展至树型网络拓扑,并在文献[9]中,提出了拓扑独立的 DANSE 方法,该方法不再受网络拓扑的约束,使 DANSE 方法可以应用于任意拓扑的网络. 为了将上述方法应用于分布式语音增强领域,基于 DANSE 方法和线性约束最小方差(linearly constrained minimum Variance, LCMV)算法,文献[13]提出了线性约束的 DANSE 方法,该线性约束为节点特定,即每个节点的约束条件可以不一致. 该方法考虑了导向矢量已知和未知的情况,并验证了所提方法的收敛性. 同样基于 DANSE 方法和 LCMV 方法,文献[14]提出了分布式的 LCMV 方法,与线性约束 DANSE 方法不同,该方法中每个节点获取的信号估计均相同,且每个节点的压缩信号为单通道信号. 上述两种方法都基于 LCMV 方法,但进行了不同的问题描述和算法设计. 文献[15]提出了基于广义特征值分解的 DANSE 方法,该方法利用期望声源的部分先验信息实现了分布式语音增强. 上述所有基于 DANSE 的方法虽然在理论上可以证明输出结果可收敛至集中式解,但是在这些算法应用于分布式语音增强时,通常无法获得与集中式一致的性能. 此外,这些方法都是根据 WASN 中节点为传声器阵列的情况进行设计,无法应用于节点为单传声器的 WASN.

针对节点是单个传声器的 WASN,学者们基于分布式数据融合技术提出了一些分布式语音增强方法,这些数据融合技术包括平均一致性算法<sup>[16]</sup>、gossip 算法<sup>[17]</sup>、diffusion 方法<sup>[18]</sup>等. 在任意网络拓扑中,文献[19]提出了基于 gossip 算法的分布式延迟求和方法,该方法在已知声学传递函数的情况下,可以获得一定的性能增益. 然而由于缺少先验信息,在实际情况下很难利用该方法实现分布式语音增强. 同理,基于 diffusion 方法提出的分布式最小方差无失真响应方法同样存在该问题<sup>[20]</sup>. 为了避免使用声学传递函数,文献[21]基于平均一致性方法提出了分布式参数化维纳滤波算法,尽管该方法利用平均一致性使每个节点获得集中式解,但是在每一次得到输出结果时,均需要节点之间进行多次数据交换. 此外,虽然该方法避免使用声学传递函数,但是其需要节点之间的延迟信息才能完成分布式语音增强. 上述针对节点单传声器提出的分布式语音增强方法存在两个弊端,分别是须要提前获取部分先验信息以及节点之间须进行多次迭代.

针对上述应用于节点为单传声器的分布式语音增强方法所存在的问题,提出一种基于语音失真加权维纳滤波的分布式语音增强方法. 首先,在全连接的无线声学传感器网络中,基于最大输入信噪比节点,提出了语音失真加权分布式维纳滤波(speech distortion weighted distributed Wiener filter, SDW-DWF)算法. 该方法由获得的最大输入信噪比节点接收其他节点发送的多帧信号信息,来分布式实现加权维纳滤波语音增强. 该算法在节点间仅须交换少量数据,且不依赖于声学传递函数. 此外,为了使所提 SDW-DWF 方法不受网络拓扑的约束,提出了一种无拓扑约束的比较一致性算法. 基于所提比较一致性算法获得具有最大输入信噪比的节点,SDW-DWF 方法即可适用于无网络拓扑约束的移动说话人场景.

## 1 信号模型

研究的分布式语音增强方法应用于每个节点为单传声器的 WASN. 考虑该场景包含  $J$  个节点且仅有一个语音源,第  $j$  个节点接收的带噪语音信号可以在短时傅里叶变换(short-time Fourier transform, STFT)域<sup>[22]</sup>表示为

$$y_j(k, l) = x_j(k, l) + v_j(k, l),$$

式中:  $x_j(k, l)$  和  $v_j(k, l)$  分别为语音信号和噪声信号,且假设二者互不相关;  $k$  和  $l$  分别为频率索引

引和时间帧索引. 在第 $k$ 个频点, 将第 $j$ 个节点的前 $L$ 帧STFT系数堆叠表示为

$$\mathbf{y}_j(k, l) = [y_j(k, l), y_j(k, l-1), \dots, y_j(k, l-L+1)]^T. \quad (1)$$

根据式(1), WASN中所有节点的前 $L$ 帧STFT系数向量可堆叠成 $JL$ 维向量, 其表示为

$$\mathbf{y}(k, l) = [\mathbf{y}_1^T(k, l), \mathbf{y}_2^T(k, l), \dots, \mathbf{y}_J^T(k, l)]^T. \quad (2)$$

同理,  $\mathbf{x}(k, l)$ 和 $\mathbf{v}(k, l)$ 的表示方式与式(2)类似. 为了表示方便, 下文所述均省略索引 $k$ 和 $l$ . 由于语音和噪声二者互不相关, 因此带噪语音信号的相关矩阵可以表示为

$$\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\} = \mathbf{R}_{xx} + \mathbf{R}_{vv},$$

式中 $\mathbf{R}_{xx}$ 和 $\mathbf{R}_{vv}$ 分别表示语音和噪声的相关矩阵. 在实际应用中, 通常会采用时间递推的方式来估计带噪语音信号的相关矩阵

$$\mathbf{R}_{yy}[l] = \alpha \mathbf{R}_{yy}[l-1] + (1-\alpha) \mathbf{y}[l] \mathbf{y}[l]^H, \quad (3)$$

式中 $\alpha$ 为遗忘因子. 利用语音活动检测(voice activity detection, VAD)<sup>[23]</sup>结果, 同样可以对噪声信号的相关矩阵进行估计. 由于语音和噪声二者互不相关, 因此最终可以得到 $\mathbf{R}_{xx} = \mathbf{R}_{yy} - \mathbf{R}_{vv}$ .

在每个节点为单传声器的WASN中, 为了实现语音增强, 通常可以设计一个线性滤波器 $\mathbf{w}$ , 并作用于带噪语音信号, 进而可以将估计到的语音信号表示为

$$\hat{d} = \mathbf{w}^H \mathbf{y}. \quad (4)$$

为了提高分布式语音增强的伸缩性, 基于加权最小均方误差(minimum mean-square error, MMSE)准则构建了下述最优化问题<sup>[24]</sup>

$$\min_{\mathbf{w}} E\{|d - \mathbf{w}^H \mathbf{x}|^2\} + \beta E\{|\mathbf{w}^H \mathbf{v}|^2\}, \quad (5)$$

式中: 第一个期望项和第二个期望项分别表示语音失真和残余噪声;  $d$ 表示期望语音;  $\beta$ 为平衡参数, 该参数可以用于调节语音增强性能. 通过观察式(5)可得出, 当 $\beta = 1$ 时, 语音失真加权问题等效于MMSE; 当 $\beta > 1$ 时, 以语音失真为代价可以换取更多的噪声消除; 当 $\beta < 1$ 时, 语音失真会非常少, 但是输出的噪声残留会非常严重. 通过对式(5)进行最优化求解, 可以得到滤波器的表达式为

$$\mathbf{w} = (\mathbf{R}_{xx} + \beta \mathbf{R}_{vv})^{-1} \mathbf{R}_{xx} \mathbf{u}_1, \quad (6)$$

式中 $\mathbf{u}_1$ 表示第一个元素为1、其余元素均为0的向量. 由于参数 $\beta$ 的存在, 因此式(6)称为语音失真加权维纳滤波. 该滤波器在频点 $k$ 仅包含前 $L$ 帧信息, 且无须未来帧信息, 因此它是因果滤波器.

根据式(6)可知: 如果将滤波器直接应用于WASN中, 每个节点不仅须要将 $L$ 维的信号发送给

其他节点, 而且实现滤波器更新的节点还须要承担 $JL$ 大小的矩阵求逆运算, 这须要占用大量的通信带宽和消耗高昂的算力. 下文将介绍分布式语音增强方法, 与集中式方法相比, 该方法可以减少数据传输所需的通信带宽, 并能实现较低维度的求逆运算, 进而减少所需的计算复杂度.

## 2 分布式语音增强方法

### 2.1 全连接网络拓扑的实现

针对全连接WASN, 将提到的语音失真加权维纳滤波方法分布式实现, 该方法可以利用语音多帧信号间的相关性, 无须任何先验信息的情况下, 提升分布式语音增强性能. 此外, 该方法基于节点特定估计方法实现分布式计算, 因此对输入信号的每一次计算, 节点间只需交换少量数据.

在WASN中, 考虑仅有的语音源处于运动状态. 随着语音源的运动, 每个节点接收的语音信号的信噪比会实时发生变化, 因此将第 $j$ 个节点当前第 $l$ 帧的STFT域输入信噪比定义为<sup>[25]</sup>

$$\mathfrak{S}_j^l = 10 \log_{10} \left[ \frac{\sum_{k=0}^{K-1} |x_j|^2}{\left( \frac{1}{l_v} \sum_{l=0}^{l-1} \sum_{k=0}^{K-1} |v_j|^2 \right)} \right], \quad (7)$$

式中: 分子项为当前帧语音信号的能量; 分母项为截止到当前帧的所有噪声帧的平均能量;  $l_v$ 为截止到当前帧所包含的噪声帧的总帧数;  $K$ 为快速傅里叶变换(fast Fourier transform, FFT)的大小. 式(7)的计算须要先依据VAD结果, 来判断当前帧是否为语音帧, 若当前帧是语音帧, 则通过分子项计算当前帧的语音信号能量, 而分母是之前帧计算得到的结果, 这样就可以在语音帧更新输入信噪比; 若当前帧是噪声帧, 则不更新输入信噪比, 而是通过分母项, 利用当前噪声帧能量, 来更新截止到当前帧的所有噪声帧的平均能量.

在WASN中实现语音增强时, 可以将估计到的语音信号表示为所有节点共同作用的结果, 因此式(4)可以重新表示为

$$\hat{d} = \sum_{j \in C} \hat{\mathbf{w}}_j^H \mathbf{y}_j, \quad (8)$$

式中:  $C = \{1, 2, \dots, J\}$ 为节点索引集合,  $\hat{\mathbf{w}}_j$ 为 $\hat{\mathbf{w}}$ 滤波器的一部分, 该部分滤波器作用的是节点 $j$ 接收到的语音信号, 称之为节点 $j$ 的本地滤波器. 通过观察式(8)可知,  $\hat{\mathbf{w}}_j$ 不仅可以作为滤波器, 也可以作为信号压缩器. 此外, 由于分布式方法须要通过迭代的方式完成, 因此后续的符号表示将会在右上角标的位置添加一个迭代索引 $i$ . 综上所述,

每一个节点压缩信号的过程可以表示为

$$z_j^i = (\hat{w}_j^i)^H y_j, \quad (9)$$

式中  $y_j$  为节点  $j$  单传声器接收到的前  $L$  帧信号。

当每一个节点均接收到来自其他所有节点发送的如式(9)所示的压缩信号后, 每一个节点就可以得到最终的分布式语音增强结果  $\hat{d} = \sum_{j \in C} z_j^i$ 。

首先, 将所有节点的压缩信号表示为

$$z^i = [z_1^i, z_2^i, \dots, z_J^i]^T.$$

此外, 在  $z^i$  中将第  $j$  个节点的压缩信号  $z_j^i$  移除后, 可以表示为  $z_{-j}^i$ 。然后, 每个节点在接收到来自其他所有节点发送的压缩信号后, 其本地的  $L+J-1$  维信号向量表示为  $\tilde{y}_j^i = [y_j; z_{-j}^i] = \tilde{x}_j^i + \tilde{v}_j^i$ 。最后, 在每一次迭代计算中, 根据式(6)及式(7)计算出的节点输入信噪比, 令具有最大输入信噪比的节点  $q$  更新其节点特定滤波器, 该滤波器的更新方式如下

$$\tilde{w}_q^{i+1} = [\hat{w}_q^{i+1}; \hat{g}_q^{i+1}] = (R_{\tilde{y}_q \tilde{y}_q}^i + \beta R_{\tilde{v}_q \tilde{v}_q}^i)^{-1} R_{\tilde{y}_q \tilde{y}_q}^i u_1, \quad (10)$$

式中:  $\hat{g}_q^{i+1} = [\hat{g}_1^{i+1}, \dots, \hat{g}_{q-1}^{i+1}, \hat{g}_{q+1}^{i+1}, \dots, \hat{g}_J^{i+1}]^T$  为作用于  $z_{-q}^i$  的滤波器;  $R_{\tilde{y}_q \tilde{y}_q}^i = E[\tilde{y}_q^i (\tilde{y}_q^i)^H]$  为向量  $\tilde{y}_q^i$  的相关矩阵,  $R_{\tilde{x}_q \tilde{x}_q}^i$  和  $R_{\tilde{v}_q \tilde{v}_q}^i$  的定义方式与  $R_{\tilde{y}_q \tilde{y}_q}^i$  相似。由于节点  $q$  接收到其他所有节点发送的压缩信号后, 可以构建本地  $L+J-1$  维信号向量  $\tilde{y}_q^i$ , 因此式(10)中的相关矩阵逆运算可以在节点  $q$  直接计算, 其所涉及相关矩阵估计方法可以参照式(3)。在  $z_{-q}^i$  中的每一个元素值为节点压缩信号  $z_j^i, j \in C \setminus \{q\}$ , 且如式(9)所示, 该压缩信号由节点本地滤波器完成, 所以通过滤波器  $\hat{g}_q^{i+1}$  作用于  $z_{-q}^i$ , 即  $(\hat{g}_q^{i+1})^H z_{-q}^i$ , 那么其他节点  $j \in C \setminus \{q\}$  就可以按照下述方式更新各自的节点本地滤波器

$$\hat{w}_j^{i+1} = \hat{g}_j^{i+1} \hat{w}_j^i. \quad (11)$$

将上述的分布式滤波器更新方法称为语音失真加权分布式维纳滤波(speech distortion weighted distributed Wiener filter, SDW-DWF)算法, 该方法的具体步骤展示在算法 1 中, 其中  $K$  为 FFT 大小。所提 SDW-DWF 方法的每一次迭代, 均执行在不同的信号帧上, 这意味着每个节点仅会将接收到的语音信号压缩并发送一次。因此, 在每一次迭代, 不会重复计算 SDW-DWF 方法的输出。另外, 在算法运行中, 输入信噪比的计算仅对语音帧有效, 当处于噪声帧时, 最大输入信噪比的节点索引保持上次迭代的值不变。由于一般情况下语音信号的初始帧均为噪声帧, 因此在算法初始化时, 将最大输入信噪比的节点索引进行随机化处理。

**算法 1** 语音失真加权分布式维纳滤波算法

**初始化**  $i \leftarrow 0, q$  为节点索引集合  $C$  中的任意元素,  $\hat{w}_j^0$  为随机向量,  $\forall j \in C$ 。

**步骤 1** 每个节点将接收到的前  $L$  帧信号  $y_j[iK+k]$  压缩为  $z_j^i[iK+k]$  ( $k=0, 1, \dots, K-1$ );

**步骤 2** 每个节点利用式(7)计算输入信噪比;

**步骤 3** 每个节点发送压缩信号和输入信噪比至其他所有节点;

**步骤 4** 每个节点生成相同的 SDW-DWF 输出信号  $\hat{d}[iK+k] = \sum_{j \in C} z_j^i[iK+k]$ ;

**步骤 5** 每个节点基于其他所有节点发送的压缩信号以及自己接收的语音信号, 更新  $L+J-1$  维向量  $\tilde{y}_j^i[iK+k]$ ;

**步骤 6** 每个节点更新相关矩阵  $R_{\tilde{y}_j \tilde{y}_j}^i$  和  $R_{\tilde{v}_j \tilde{v}_j}^i$ ;

**步骤 7** 节点  $q$  通过式(10)计算本地滤波器  $\tilde{w}_q^{i+1}$ , 并将向量  $\hat{g}_q^{i+1}$  发送给其他所有节点;

**步骤 8** 其他节点  $j \in C \setminus \{q\}$  通过式(11)更新  $\hat{w}_j^{i+1}$ ;

**步骤 9** IF 该帧信号为噪声帧 THEN;

**步骤 10** 不更新索引值  $q$ ;

**步骤 11** ELSE;

**步骤 12** 每个节点和其他节点发送的输入信噪比进行比较, 最终将  $q$  更新为具有最大输入信噪比的节点索引;

**步骤 13** END IF;

**步骤 14**  $i = i + 1$ ;

**步骤 15** 返回步骤 1。

## 2.2 无网络拓扑约束的实现

所提的 SDW-DWF 算法只能应用于全连接 WASN, 该网络结构意味着每个节点能够和其他所有节点进行数据传输。当不指定网络结构, 即节点之间的连接拓扑由节点自组织生成时, WASN 中节点间生成的网络拓扑为任意结构。针对无网络拓扑约束的场景, 提出了一种比较一致性算法, 该方法通过令每一个节点和其邻居节点经过数次数据传输后, 所有节点均能获得整个 WASN 中的最大输入信噪比值。首先定义一个连接矩阵  $L^{[21]}$

$$L_{jq} = \begin{cases} 1 & (q \in N_j); \\ 1 & (j = q); \\ 0 & (\text{其他}), \end{cases}$$

式中:  $L_{jq}$  为连接矩阵中第  $j$  行第  $q$  列的元素;  $N_j$  为节点  $j$  邻居节点的集合, 且不包含它本身。然后, 将每一次迭代中各个节点计算得到的输入信噪比堆

叠到一个  $J \times J$  矩阵中, 该矩阵表示为

$$\mathbf{S}^i = \begin{bmatrix} \mathfrak{S}_1^i & \mathfrak{S}_2^i & \cdots & \mathfrak{S}_J^i \\ \mathfrak{S}_1^i & \mathfrak{S}_2^i & \cdots & \mathfrak{S}_J^i \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{S}_1^i & \mathfrak{S}_2^i & \cdots & \mathfrak{S}_J^i \end{bmatrix},$$

式中: 输入信噪比的右上角标索引更换为了迭代索引, 因为一次迭代等效于一帧, 所以这两个索引具备相同的含义. 当每一个节点将输入信噪比发送给邻居节点后, 可以用  $\mathbf{L} \odot \mathbf{S}^i$  表示每一个节点获取到的输入信噪比, 其中  $\odot$  表示哈达玛积, 此时每一个节点将自己的输入信噪比和邻居节点输入信噪比进行比较, 并将最大的输入信噪比更新为自己的输入信噪比值. 最终, 经过数次节点输入信噪比的发送比较, 每个节点均能获得整个 WASN 中的最大输入信噪比值. 该方法所需的节点传输次数取决于 WASN 中的节点数, 即节点数越多, 需要传输的次数越多.

结合所提比较一致性算法, 将提出的 SDW-DWF 算法应用于无拓扑约束的 WASN, 具体步骤呈现在算法 2. 算法 2 与算法 1 的不同点在于更新节点的本地向量构建方式, 以及最大输入信噪比的比较方式, 而其余步骤具有相似性.

**算法 2** 结合比较一致性算法的语音失真加权分布式维纳滤波算法

**初始化**  $i \leftarrow 0$ ,  $q$  为节点索引集合  $C$  中的任意元素,  $\hat{\mathbf{w}}_q^0$  为随机向量,  $\forall j \in C$ .

**步骤 1** 节点  $q$  的所有邻居节点将接收到的前  $L$  帧信号  $\mathbf{y}_{N_q}[iK+k]$  压缩为  $\mathbf{z}_{N_q}^i[iK+k]$  ( $k=0, 1, \dots, K-1$ ), 并发送给节点  $q$ ;

**步骤 2** 每个节点利用式(7)计算输入信噪比;

**步骤 3** 通过使用比较一致性算法, 使得每个节点均获得最大输入信噪比的值;

**步骤 4** 节点  $q$  生成 SDW-DWF 输出信号

$$\hat{\mathbf{d}}[iK+k] = (\hat{\mathbf{w}}_q^i)^H \mathbf{y}_q[iK+k] + \sum_{j \in N_q} \mathbf{z}_j^i[iK+k];$$

**步骤 5** 节点  $q$  基于邻居节点发送的压缩信号以及自己接收的语音信号, 更新堆叠向量  $\hat{\mathbf{y}}_q^i[iK+k]$ , 剩余的其他所有节点同样基于接收的语音信号更新向量  $\hat{\mathbf{y}}_j^i[iK+k]$ , 其中压缩信号均为 0 元素, 且压缩信号的个数取决于邻居节点数;

**步骤 6** 每个节点更新相关矩阵  $\mathbf{R}_{\hat{\mathbf{y}}_j^i}^i$  和  $\mathbf{R}_{\hat{\mathbf{y}}_q^i}^i$ ;

**步骤 7** 节点  $q$  通过式(10)计算本地滤波器  $\hat{\mathbf{w}}_q^{i+1}$ , 并将向量  $\hat{\mathbf{g}}_q^{i+1}$  发送给其所有邻居节点, 并通过式(11)更新  $\hat{\mathbf{w}}_{N_q}^{i+1}$ ;

**步骤 8** 节点  $q$  利用连接矩阵将 SDW-DWF 的输出信号扩散至所有节点;

**步骤 9** IF 该帧信号为噪声帧 THEN;

**步骤 10** 不更新索引值  $q$ ;

**步骤 11** ELSE;

**步骤 12** 每个节点和最大输入信噪比进行比较, 最终将  $q$  更新为具有最大输入信噪比的节点索引;

**步骤 13** END IF;

**步骤 14**  $i = i + 1$ ;

**步骤 15** 返回步骤 1.

### 2.3 数据传输与计算复杂度分析

为了便于分析数据传输, 假设帧长与 FFT 大小相等, 且帧之间不存在重叠部分, 该假设只适用于本小节. 针对算法 1, WASN 中的每个节点均须要发送输入信噪比和压缩信号, 因此整个网络每秒须要发送的采样点数为  $J(f_s/K + f_s)$ , 其中  $f_s$  表示信号采样频率. 另外, 由于更新滤波器的节点还须要将对应压缩信号的滤波器发送至其余所有节点, 所以每秒还须要发送额外的  $(J-1)f_s$  个采样. 综上所述, 算法 1 在整个网络中每秒一共须要传输的采样点数为  $(J/K + 2J-1)f_s$ .

在算法 2 中, 设比较一致性算法须要节点间传输  $T$  次, 能使每个节点获得最大输入信噪比, 因此关于输入信噪比的传输, 整个 WASN 每秒须要发送的采样点数为  $JTf_s/K$ . 由于更新节点须要将对应压缩信号的滤波器发送给邻居节点, 且邻居节点还须要发送压缩信号给更新节点, 因此每秒还须要发送的采样点数为  $2|N_q|f_s$ , 其中  $|N_q|$  为更新节点的邻居节点数. 此外, 将增强信号扩散至所有节点最多须要发送  $Jf_s$  个采样点. 综上可见, 算法 2 在整个网络中每秒一共须要传输的采样点数为  $(JT/K + 2|N_q| + J)f_s$ .

当分析计算复杂度时, 设  $M \times M$  矩阵求逆运算的计算复杂度为  $O(M^3)^{[5]}$ . 在算法 1 中, 更新节点处的矩阵大小为  $(L+J-1) \times (L+J-1)$ , 所以该矩阵求逆导致所需的计算复杂度为  $O(L+J-1)^3$ . 同理, 算法 2 所需的计算复杂度为  $O(L+|N_q|)^3$ . 在本文方法对应的集中式方法中, 求逆矩阵的大小为  $JL \times JL$ . 当 WASN 中的节点数量固定时, 随着帧数的不断增大, 分布式方法的计算复杂度要远小于集中式方法, 因为在分布式方法中, 仅有一个节点获取到多帧信号, 其余信号均来自其他节点压缩后的单通道信号.

### 3 仿真实验

#### 3.1 实验设置

在语音信号处理中,混响时间是一个非常重要的参数,为了准确评估语音信号处理算法在混响条件下的性能,往往需要一个能准确生成混响时间的仿真模型.因此,采用了 Image-Source 方法<sup>[26]</sup>仿真了一个大小为  $7\text{ m}\times 7\text{ m}\times 3\text{ m}$  的房间,并在该房间中设置了 7 个随机分布的节点,每个节点为单传声器.节点分布以及节点间的连接拓扑如图 1 所示,其中  $X$  轴和  $Y$  轴分别表示房间的长和宽.同时,在该房间中仿真了一个运动说话人情景,该说话人的起始位置为  $(1\text{ m}, 1\text{ m})$ ,中间途经  $(4.5\text{ m}, 1\text{ m})$ ,终点为  $(4.5\text{ m}, 4\text{ m})$ .说话人从起始位置匀速直线运动至途经点总耗时为  $3.2\text{ s}$ ,再从途经点匀速直线运动至终点总耗时为  $2.8\text{ s}$ ,设置全程坐标的变化间隔为  $0.02\text{ s}$ .此外,设置所有的节点高度为  $1\text{ m}$ ,说话人高度为  $1.7\text{ m}$ ,声音在室内的传播速度为  $340\text{ m/s}$ ,采样频率为  $16\text{ kHz}$ .由上述提供的房间大小、节点位置、说话人移动轨迹、声速及采样频率等信息,生成了混响时间为  $0.5\text{ s}$  的房间冲激响应.从 TIMIT 数据库<sup>[27]</sup>中随机抽取 10 段语音,分别作为实验期间的声源信号,其采样频率为  $16\text{ kHz}$ ,且时长统一为  $6\text{ s}$ ,所有实验结果均为 10 段语音的平均值.当声源信号和各节点对应的房间冲激响应卷积后,令混响语音信号与来自 NOISEX-92 数据库<sup>[28]</sup>的噪声相加,最终生成具备指定信噪比的带噪语音信号,其中白噪声和 babble 噪声将分别作为背景噪声进行算法验证,其采样频率下采样为  $16\text{ kHz}$ .

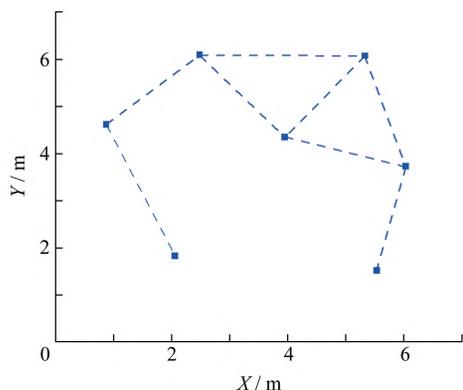


图1 无线声学传感器网络拓扑示意图

在实现 STFT 时,设置帧长为 256 点,帧之间的重叠率为 75%,FFT 大小为 1 024 点,分析窗和合成窗分别采用汉宁窗函数和海明窗函数.此外,

本文实验采用时间递推的方式估计信号相关矩阵,根据经验,遗忘因子大小设置为 0.997.

#### 3.2 评价指标

实验期间,评价实验结果的指标分别采用语音质量感知评估(perceptual evaluation of speech quality, PESQ)和短时客观可懂度(short-time objective intelligibility, STOI)<sup>[29]</sup>.PESQ 是国际电信联盟电信标准化部 (ITU-T) P.862 建议书提供的语音质量评价指标,它的评分范围为  $-0.5\sim 4.5$ ,其值越大,说明语音质量越好.STOI 的评分值与语音可懂度呈递增关系,其评分范围是从  $0\sim 1$ .

#### 3.3 实验结果与比较

为了验证本文方法的有效性,假设图 1 所示的节点连接拓扑为全连接,分别验证了本文方法涉及的两个不同参数对其分布式语音增强性能的影响.其次,在如图 1 所示的节点连接拓扑下,本文方法分别与文献[19]提出的分布式延迟求和波束形成(distributed delay and sum beamformer, DDSB)方法,以及文献[21]提出的分布式参数化多通道维纳滤波(distributed parameterized multichannel Wiener filter, DPMWF)方法和简化的分布式参数化多通道维纳滤波(simplified distributed parameterized multichannel Wiener filter, SDPMWF)方法进行了比较,其中本文方法涉及的帧数设置为 3,平衡参数设置为 15,DPMWF 和 SDPMWF 中涉及的平衡参数也设置为 15.此外,DDSB, DPMWF 和 SDPMWF 都必须提前获取声学传递函数,所以在实现这三种对比方法时,采取离线估计的方式先行获得声学传递函数,再完成这三种滤波器的分布式估计.最后,在会议室录制了语音,并利用上述的分布式算法分别对录音结果进行增强,比较了各方法在实际场景中的性能.

##### 3.3.1 不同帧数对所提方法性能的影响

本实验分别在白噪声和 babble 噪声下,利用 PESQ 评价指标测试不同帧数对所提方法性能的影响,其中所提方法涉及的平衡参数设置为 15,帧数范围为  $2\sim 10$ ,步长为 1.针对每一种噪声,均设置不同的输入信噪比,分别为  $-5, 0, 5, 10$  和  $15\text{ dB}$ .图 2 和图 3 分别展示了白噪声和 babble 噪声作为背景噪声时,提出的 SDW-DWF 方法增强后语音以及节点接收带噪语音的 PESQ 指标变化情况,其中:实线表示 SDW-DWF 方法;虚线表示带噪语音.此外,因为整个 WASN 中有 7 个节点,所以带噪语音的 PESQ 值为 7 个节点接收带噪语音信号的 PESQ 平均值,而在实现提出的分布式语音增强方

法时, 每个节点获取的增强信号完全相同, 因此图中展示的SDW-DWF性能为单个节点增强后结果, 也可以理解为所有节点的平均结果。

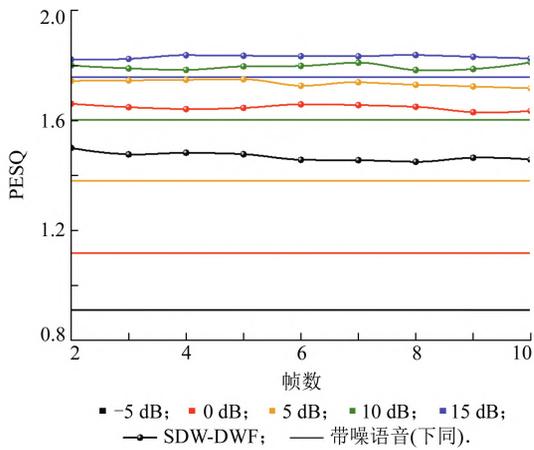


图2 不同输入信噪比的白噪声下,PESQ随帧数的变化

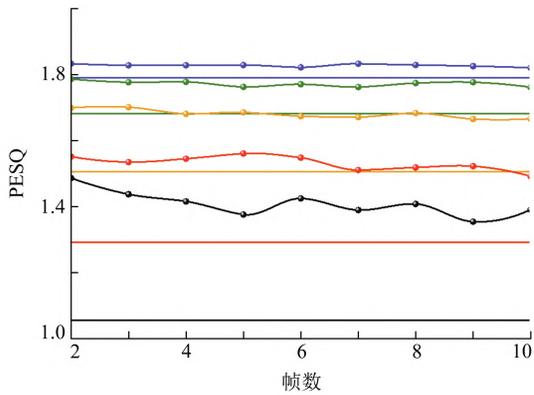


图3 不同输入信噪比的babble噪声下,PESQ随帧数的变化

由图2和图3可知: 不论是带噪语音的PESQ, 还是SDW-DWF的PESQ, 均会随着输入信噪比的升高而增加. 另外, 随着输入信噪比的增加, SDW-DWF方法相对于带噪语音的PESQ增益变得越来越小, 不过无论输入信噪比多少, 本文方法均能在两种噪声下实现一定的分布式语音增强性能. 这得益于本文方法利用了多帧信号间的相关性, 并且将具有最大输入信噪比的节点作为参考节点. 此外, 通过观察图2和图3, 在低输入信噪比下, 所提SDW-DWF方法的帧数为2或3, 带来的性能提升比较突出, 而在高输入信噪比时, 本文方法的帧数为8, 带来的性能也较为明显, 这说明提出分布式方法针对低信噪比的输入信号, 利用少数帧的相关性即可完成更好的噪声消除, 而针对高信噪比的输入信号, 需要更多帧的数据相关性来实现最佳的增强效果。

### 3.3.2 不同参数值对所提方法性能的影响

分别在白噪声和babble噪声下利用PESQ评价指标测试不同平衡参数值对所提方法性能的影响,

其中本文方法涉及的帧数设置为3, 平衡参数范围为1~25, 步长为1. 该实验的其他所有配置均与3.3.1节实验相同. 图4和图5分别展示了白噪声和babble噪声作为背景噪声时, 所提的SDW-DWF方法增强后语音以及节点接收带噪语音的PESQ指标变化情况。

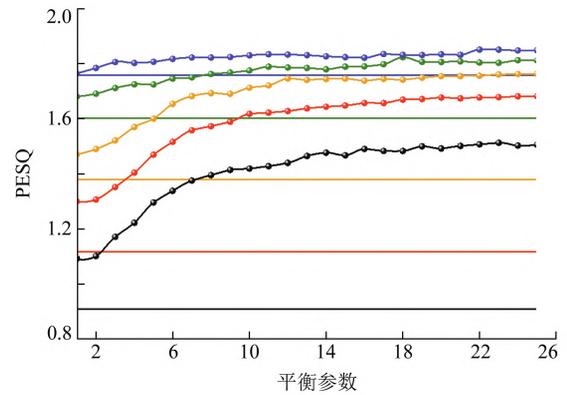


图4 不同输入信噪比的白噪声下,PESQ随平衡参数的变化

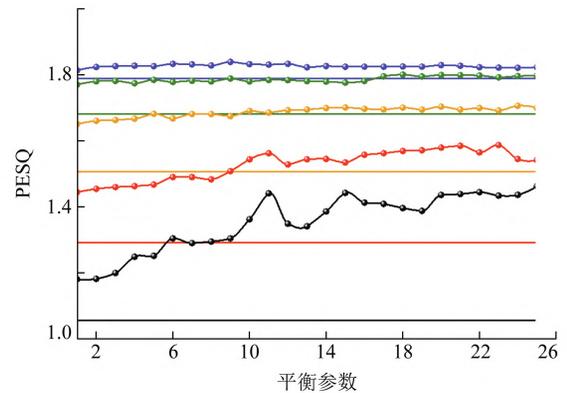


图5 不同输入信噪比的babble噪声下,PESQ随平衡参数的变化

由图4可知: 无论白噪声作为背景噪声的输入信噪比高低, 本文方法的性能均随着平衡参数的增加而提高, 当平衡参数大于15后, 本文方法的性能处于缓慢增长的趋势. 这是由于平衡参数值越大, 噪声消除量越大的同时, 造成的语音失真也越大, 因此导致性能提升逐步变小; 同样, 随着输入信噪比的增加, 所提的SDW-DWF方法相对于带噪语音的PESQ增益越来越小. 由图5可知: 当背景噪声为babble噪声且输入信噪比为-5 dB时, 本文方法的性能随着平衡参数的增加出现了较大波动, 但整体上仍然呈现递增关系, 且具有较高的性能提升. 在其他输入信噪比情况下, 本文方法的性能均随着平衡参数的增加而提高, 当平衡参数大于15后, 本文方法的性能趋于稳定. 该背景噪声的结果与白噪声作为背景噪声时有所不同, 因为babble噪声属于非平稳噪声, 较白噪声而言, 更难消除, 导

致本文方法的平衡参数值在 15 左右时, 性能提升就达到上限. 总体而言, 不论在哪种背景噪声下, 且无论输入信噪比高低, 本文方法的性能均能由平衡参数实现调节.

### 3.3.3 不同方法的性能比较

本实验采用 PESQ 和 STOI 两个评价指标对比较方法进行测试, 其中给出的测试结果为每个节点评价值的平均结果. 实验期间, 分别采用白噪声和 babble 噪声作为背景噪声, 且针对每种噪声, 均设置不同的输入信噪比, 分别为 -5, 0, 5, 10 和 15 dB.

表 1~4 展示了不同背景噪声下对应的 PESQ 和 STOI 评价结果, 其中除了比较四种方法在自组织拓扑中的性能外, 为说明本文方法的必要性, 表中还包含有 7 个节点分别对应的单通道算法性能的平均值.

表 1 不同输入信噪比的白噪声下, 自组织拓扑中方法的 PESQ 比较

输入信噪比/dB	节点	DDSB	DPMWF	SDPMWF	SDW-DWF	单传声器
-5	0.909	1.028	0.999	1.009	1.455	1.457
0	1.117	1.186	1.262	1.239	1.633	1.605
5	1.380	1.434	1.413	1.457	1.737	1.695
10	1.601	1.623	1.656	1.680	1.797	1.764
15	1.758	1.808	1.786	1.816	1.823	1.811

表 2 不同输入信噪比的白噪声下, 自组织拓扑中方法的 STOI 比较

输入信噪比/dB	节点	DDSB	DPMWF	SDPMWF	SDW-DWF	单传声器
-5	0.467	0.451	0.453	0.457	0.470	0.406
0	0.510	0.511	0.501	0.511	0.531	0.451
5	0.548	0.532	0.553	0.569	0.593	0.495
10	0.573	0.579	0.607	0.623	0.641	0.530
15	0.587	0.616	0.649	0.662	0.673	0.557

表 3 不同输入信噪比的 babble 噪声下, 自组织拓扑中方法的 PESQ 比较

输入信噪比/dB	节点	DDSB	DPMWF	SDPMWF	SDW-DWF	单传声器
-5	1.055	1.165	1.102	1.037	1.348	1.419
0	1.292	1.369	1.317	1.353	1.534	1.536
5	1.507	1.577	1.556	1.581	1.667	1.644
10	1.682	1.737	1.764	1.776	1.789	1.730
15	1.791	1.814	1.878	1.924	1.924	1.804

表 4 不同输入信噪比的 babble 噪声下, 自组织拓扑中方法的 STOI 比较

输入信噪比/dB	节点	DDSB	DPMWF	SDPMWF	SDW-DWF	单传声器
-5	0.443	0.451	0.455	0.461	0.479	0.419
0	0.486	0.483	0.492	0.501	0.541	0.451
5	0.522	0.524	0.538	0.555	0.588	0.485
10	0.549	0.568	0.595	0.614	0.634	0.519
15	0.568	0.600	0.635	0.655	0.663	0.547

表 1 结果表明: 在白噪声背景下, 无论输入信噪比为多少, 对比的三种方法所带来的 PESQ 提升有限, 而本文方法展现了对 PESQ 测试指标较为明显的提升. 另外, 与单传声器方法相比, 本文方法只有在输入信噪比为 -5 dB 时, 略微低于单传声器的 PESQ 值. 除此之外, 所有条件下的本文方法测试结果, 均可以证明分布式方法的优势, 所提分布式方法不仅可以利用多个节点进行语音增强, 还可以节省不必要的数据通信和算力. 证实该方法的有效性后, 可以使分布式语音增强技术应用于更广泛的场景. 在表 2 中, 通过观察各方法的 STOI 值, 可以看出本文方法不仅实现了 STOI 提升, 而且提升值与输入信噪比呈递增关系, 该趋势与表 1 结果相反. 这是由于 PESQ 测试的是语音质量, 在低输入信噪比时, 本文方法可以消除较多的噪声量, 所以提升语音质量较为明显, 尽管如此, 语音可懂度依然很低, 因此 STOI 提升值较低. 此外, 单传声器方法的结果展示了其无法提升语音可懂度, 该结果再一次验证了本文方法的必要性与可行性. 此外, 在 babble 噪声下, 表 3 和表 4 的结果同样可以得出与表 1 和表 2 类似的结论. 综上所述, 本文方法在无须声学传递函数的前提下, 就能获得较好的分布式语音增强性能.

### 3.3.4 实际录音测试

为了验证本文方法的有效性与可行性, 在会议室进行了实际录音测试. 本实验在一个 8 m×10 m×2.5 m 的会议室中, 进行实际录音, 该录音环境如图 6 所示. 具体地, 会议桌上随机摆放着 6 部手机负责录制声音, 声源则是由音箱播放一段时长为 36 s 的纯净语音, 噪声源为平板播放的 babble 噪声, 背景噪声为真实环境噪声. 在图 6 中, 点划线表示 6 个节点之间的连接拓扑. 该实验是在声音录制结束后, 将 6 个节点录制的语音信号进行算法仿真测试.

因为实际环境中录制的声音要比仿真生成的带

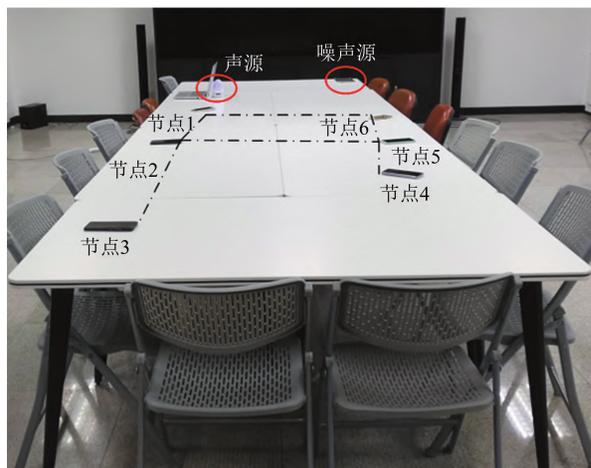


图6 实际录音环境

噪语音复杂, 且两种情况下的语音时长不相同, 所以本文方法在这次实验中, 再次对不同帧数进行了测试, 帧数范围为3~27, 步长为3. 通过比较实验结果发现, 帧数为18时, PESQ达到峰值, 而STOI值与步长为单调递增关系, 所以在表5中展示了帧数为18时的所提方法性能. 其他方法的设置同3.3.3节实验.

表5展示了PESQ和STOI两个评价指标对所有方法的测试结果, 其中将6个节点采集的带噪语音信号性能进行平均. 从表5可以看出: DDSB方法没有提升PESQ和STOI, 而DPMWF和SDPMWF方法略微提升, 提升效果最为明显的是本文方法. 尽管对比方法采用了声学传递函数的离线估计, 但由于其过度依赖声学传递函数, 因此受声学传递函数估计误差的影响, 这些方法带来的性能提升并不明显. 而本文方法无须额外的先验信息, 仅利用多帧之间的相关性, 即可对实际环境录制的语音实现较高的性能提升.

表5 实际录音测试中的性能比较

方法	PESQ	STOI
节点	1.607	0.391
DDSB	1.607	0.391
DPMWF	1.637	0.392
SDPMWF	1.630	0.392
SDW-DWF(L=18)	1.767	0.417

## 4 结语

在节点为单传声器的无线声学传感器网络中, 提出了一种基于语音失真加权维纳滤波的分布式语音增强算法. 首先, 基于全连接WASN的最大输入信噪比节点, 提出了语音失真加权分布式维纳滤

波算法. 该方法通过最大输入信噪比节点接收其它节点发送的多帧信号信息, 来分布式实现加权维纳滤波语音增强. 该算法在节点间仅需交换少量数据, 且不依赖于声学传递函数. 此外, 还提出了无拓扑约束的比较一致性算法, 该方法使得所提分布式算法可用于无网络拓扑约束的移动说话人场景. 最后, 仿真和实际实验验证了本文方法的有效性与必要性.

## 参 考 文 献

- [1] 贾海蓉, 梅淑琳, 张敏. 基于双通道神经网络时频掩蔽的语音增强算法[J]. 华中科技大学学报(自然科学版), 2021, 49(6): 43-49.
- [2] 李如玮, 李秋艳, 赵丰年, 等. 基于注意力和深度学习的双耳语音增强算法[J]. 华中科技大学学报(自然科学版), 2023, 51(9): 125-131.
- [3] ZHANG Qiquan, QIAN Xinyuan, NI Zhaoheng, et al. A time-frequency attention module for neural speech enhancement[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 462-475.
- [4] LI Yihao, ZHANG Xiongwei, SUN Meng. A unified speech enhancement approach to mitigate both background noises and adversarial perturbations[J]. Information Fusion, 2023, 95: 372-383.
- [5] BERTRAND A, MOONEN M. Distributed adaptive node-specific signal estimation in fully connected sensor networks, part I: sequential node updating[J]. IEEE Transactions on Signal Processing, 2010, 58(10): 5277-5291.
- [6] BERTRAND A, MOONEN M. Distributed adaptive node-specific signal estimation in fully connected sensor networks, part II: simultaneous and asynchronous node updating[J]. IEEE Transactions on Signal Processing, 2010, 58(10): 5292-5306.
- [7] BERTRAND A, MOONEN M. Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology[J]. IEEE Transactions on Signal Processing, 2011, 59(5): 2196-2210.
- [8] ZENG Yuan, HENDRIKS R C. Distributed estimation of the inverse of the correlation matrix for privacy preserving beamforming[J]. Signal Processing, 2015, 107: 109-122.
- [9] SZURLEY J, BERTRAND A, MOONEN M. Topology-independent distributed adaptive node-specific signal estimation in wireless sensor networks[J]. IEEE Transactions on Signal and Information Processing over Networks, 2017, 3(1): 130-144.

- [10] ZHANG Jie, CHEPURI S P, HENDRIKS R C, et al. Microphone subset selection for MVDR beamformer based noise reduction[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(3): 550-563.
- [11] ZHANG Jie, KOUTROUVELIS A I, HEUSDENS R, et al. Distributed rate-constrained LCMV beamforming[J]. *IEEE Signal Processing Letters*, 2019, 26(5): 675-679.
- [12] ZHANG Jie, DU Jun, DAI Lirong. Sensor selection for relative acoustic transfer function steered linearly-constrained beamformers[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, 29: 1220-1232.
- [13] BERTRAND A, MOONEN M. Distributed node-specific LCMV beamforming in wireless sensor networks[J]. *IEEE Transactions on Signal Processing*, 2012, 60(1): 233-246.
- [14] BERTRAND A, MOONEN M. Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission[J]. *IEEE Transactions on Signal Processing*, 2013, 61(13): 3447-3459.
- [15] ROMPAEY R V, MOONEN M. Distributed adaptive signal estimation in wireless sensor networks with partial prior knowledge of the desired sources steering matrix[J]. *IEEE Transactions on Signal and Information Processing over Networks*, 2021, 7: 478-492.
- [16] LIN Xiao, BOYD S. Fast linear iterations for distributed averaging[C]// *Proc of The 42nd IEEE International Conference on Decision and Control*. Piscataway: IEEE, 2003: 4997-5002.
- [17] KEMPE D, DOBRA A, GEHRKE J. Gossip-based computation of aggregate information[C]// *Proc of The 44th Annual IEEE Symposium on Foundations of Computer Science*. Cambridge: IEEE, 2003: 482-491.
- [18] LOPES C G, SAYED A H. Diffusion least-mean squares over adaptive networks[C]// *Proc of IEEE International Conference on Acoustics, Speech and Signal Processing*. Honolulu: IEEE, 2007: 917-920.
- [19] ZENG Yuan, HENDRIKS R C. Distributed delay and sum beamformer for speech enhancement via randomized gossip[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014, 22(1): 260-273.
- [20] O'CONNOR M, KLEIJN W B. Diffusion-based distributed MVDR beamformer[C]// *Proc of IEEE International Conference on Acoustics, Speech and Signal Processing*. Florence: IEEE, 2014: 810-814.
- [21] CHANG Ruijiang, CHEN Zhe, YIN Fuliang. Distributed multichannel Wiener filtering for speech enhancement in acoustic sensor networks[J]. *International Journal of Adaptive Control and Signal Processing*, 2022, 36(11): 2732-2753.
- [22] ZHIVOMIROV H. On the development of STFT-analysis and ISTFT-synthesis routines and their practical implementation[J]. *TEM Journal*, 2019, 8(1): 56-64.
- [23] CHENG Ming, WANG Weiqing, ZHANG Yucong, et al. Target-speaker voice activity detection via sequence-to-sequence prediction[C]// *Proc of IEEE International Conference on Acoustics, Speech and Signal Processing*. Piscataway: IEEE, 2023: 1-5.
- [24] ZHANG Jie, TAO Rui, DU Jun, et al. SDW-SWF: speech distortion weighted single-channel Wiener filter for noise reduction[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023, 31: 3176-3189.
- [25] CHANG Ruijiang, CHEN Zhe, YIN Fuliang. Robust distributed noise suppression in acoustic sensor networks[J]. *IEEE Sensors Journal*, 2022, 22(18): 18151-18161.
- [26] LEHMANN E, JOHANSSON A. Prediction of energy decay in room impulse responses simulated with an image-source model[J]. *The Journal of the Acoustical Society of America*, 2008, 124(1): 269-277.
- [27] GAROFOLO J S, et al. DARPA TIMIT acoustic-phonetic speech database[J]. *National Institute of Standards and Technology*, 1988, 15: 29-50.
- [28] VVRGA A, STEENEKEN H J. Assessment for automatic speech recognition: II. NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems[J]. *Speech Communication*, 1993, 12(3): 247-251.
- [29] TAAL C H, HENDRIKS R C, HEUSDENS R, et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19(7): 2125-2136.