ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 2,Mar. 2023,pp. 283－292
DOI:10. 16337/j. 1004-9037. 2023. 02. 005

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于联合图学习的多通道语音增强方法

张鹏程[1]，郭海燕[1]，王婷婷[1]，杨　震[1,2]

(1. 南京邮电大学通信与信息工程学院,南京 210003;2. 南京邮电大学通信与网络技术国家地方联合工程研究中心,南京 210003)

摘　要：考虑到通道间存在的空间关系影响着其降噪问题,图信号处理可以捕获该潜在关系,若直接采用其空间物理分布图,无法实时反映其时变特性,因此本文提出了一种基于联合图学习的多通道语音增强方法。首先,提出一种联合时间-空间图学习方法,以最小化多通道含噪语音信号在空间图上的平滑度、参考通道信号在语音帧内图上的平滑度、空间图的稀疏度和帧内图的稀疏度之和为目标,优化阵列空间图和语音帧内图。基于学习的空间图和帧内图,构建多通道语音信号的时间-空间联合图。在此基础上,将多通道语音图信号进行联合图傅里叶变换,进而采用固定波束形成(Fixed beam forming, FBF)方法进行增强。实验结果表明,与传统的FBF方法相比,所提出的基于联合图学习的FBF(Joint graph learning based FBF, JGL-FBF)方法显著提升了增强语音的信噪比(Signal-to-noise ratio, SNR)和主观语音质量评估(Perceptual evaluation of speech quality, PESQ)。另外,实验结果也表明,JGL-FBF方法的语音增强性能会受到时延补偿准确性的影响。

关键词：联合图学习;语音增强;多通道;波束形成

中图分类号：TN911.7　　**文献标志码**:A

Multi-channel Speech Enhancement Based on Joint Graph Learning

ZHANG Pengcheng[1], GUO Haiyan[1], WANG Tingting[1], YANG Zhen[1,2]

(1. College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;2. National Local Joint Engineering Research Center for Communications and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Considering that the spatial relationship between channels affects the noise reduction, graph signal processing can capture the potential relationship. If the spatial physical distribution map is directly used, its time-varying characteristics cannot be reflected in real time. Therefore, we propose a multi-channel speech enhancement method based on joint graph learning. Firstly, we propose a joint time-space graph learning method, which jointly optimizes the array space graph and the speech frame inner graph, for the sake of minimizing the sum of the smoothness of the multi-channel noisy speech signal on the spatial graph, the smoothness of the nosiy speech signal from the reference channel on the speech frame graph, the sparsity of the Laplace matrix and the sparsity of the adjacency matrix. Based on the learned space graph and frame inner graph, the time-space joint graph of multi-channel speech signal is constructed. On this basis, the multi-channel speech graph signal is enhanced by applying the joint graph transform and the

基金项目:国家自然科学基金(62071242)。
收稿日期:2022-07-18;**修订日期**:2022-09-24

fixed beam forming (FBF) method. Experimental results show that the proposed joint graph learning based FBF (JGL-FBF) method can significantly improve the signal-to-noise ratio (SNR) of enhanced speech and perceptual evaluation of speech quality (PESQ) compared with the traditional FBF method. In addition, the experimental results also show that the accuracy of delay compensation affects the speech enhancement performance of JGL-FBF.

Key words: joint graph learning; speech enhancement; multi-channel; beam forming

引　言

现代信号处理中往往涉及处理大规模的高维度数据,比如传感器网络、社交网络、交通网络等数据类型。与传统的时域空域信号相比,这些高维信号往往具有不规则的拓扑结构,这使得传统的数字信号处理方法无法直接应用。图信号处理(Graph signal processing, GSP)作为一种全新的数据处理技术,通过加权图揭示信号间的相互作用和联系,将传统数字信号处理理论扩展到不规则的图信号上,为处理具有复杂结构的数据提供了有效的手段。图信号处理研究涉及图结构[1]、图傅里叶变换[2]、图滤波器[3]、图频率[4]、在图像处理[5]、图信号的采样与重建[6-7]、机器学习[8]、语音信号处理[9]、无线传感网[10]等领域得到了广泛应用。

信号的图结构是GSP的研究基础。在基于GSP的语音处理领域,通常采用构建静态图(即图结构是确定的)的方式来表征语音样点间的拓扑结构关系。例如,文献[11]利用联合移位算子构建语音图信号,在此基础上获取图傅里叶基进行图傅里叶变换,对选定音频片段的图傅里叶系数进行编码,实现零水印的嵌入;文献[12]分别在时域和空域定义了两张图,用多重信号分类的方法从观测数据中估计噪声子空间,实现麦克风阵列中方向角的估计;文献[13]提出利用有向循环图构建语音图信号的图拓扑结构,采用图谱减法来进行语音增强。文献[14]定义了语音图信号的无向图,并用图邻接矩阵的迹来估计噪声的功率谱密度,取得了良好的消噪效果。文献[15]考虑阵列之间位置关系和相位差关系,定义了多通道语音邻接矩阵权值,并基于图频域的MVDR波束形成方法来进行多通道语音增强。

然而,上述研究都是对拓扑结构先验已知的信号进行处理和分析,但是在实际应用中还有大量没有明显结构或者不适合自定义图的信号。为了反映信号之间的相关性,研究者早期致力于研究图模型来恢复稀疏精度矩阵[16]。Dong等[17]从广义因子分析模型的角度表示图信号,并通过图信号的平滑性和最小化图上信号的变化,提出了一个图学习的框架。Egilmez等[18]研究在给定的结构约束(如图的连通性和稀疏度)下,从观测数据中估计图的拉普拉斯矩阵。Rabbat等[19]提出了一种基于软阈值的图估计方法,并且计算得到图在稀疏图情况下重建误差的结果。Chepuri等[20]提出将图学习问题建模为稀疏边采样函数,用基于平滑度的边选择机制来表示图的拉普拉斯形式,使图学习问题得以简化。Subbareddy等[21]提出了从宽带图信号中学习图的新准则,先从数据中学习图的特征向量,然后根据估计的正交基恢复反映图拓扑的邻接矩阵。Kadambari等[22]提出了积图学习的框架,将要估计的图拉普拉斯矩阵分解为两个小因子图的笛卡尔乘积。采用上述图学习方法虽然能够解决信号拓扑结构未知的问题,但忽略了信号在时空的演变特性。针对此,Li等[23]通过将时空信号进行建模,提出了一个图信号的时间和顶点的联合图学习框架用于海面温度的图学习。具体地,文献[23]的优化函数中包括3组优化目标,降噪信号的预测、时间图和顶点图的学习,即信号去噪直接通过联合图学习得到。

针对多通道语音信号,静态GSP技术仍有一定的局限性。而基于图学习的GSP技术能有效弥补其不足,更好地适用于实时场景。此外,基于时间-空间的联合图信号处理方法能对语音图信号的频谱进行动态处理,从而能更好地分析语音的联合图频谱信息。虽然麦克风阵列有自身的空间物理分布,

但仅根据空间分布信息来构建麦克风阵列空间图无法反映实际环境中信号随噪声、声源位置的时变特性。基于此,本文研究面向多通道语音信号的联合图学习方法,构建多通道语音图信号的联合时间-空间图拓扑结构,并在此基础上进行多通道语音增强。主要创新工作如下:

(1)所提出的联合图学习方法旨在得到反映麦克风阵列接收信号相关特性的两张图,即一张时间图和一张空间图,进而进行联合图傅里叶变换去分析语音的联合图频谱信息,然后通过麦克风阵列波束形成方法对多通道语音信号进行语音增强。本文的信号去噪是利用联合图学习后的两张图在图频域进行波束形成处理得到的。

(2)提出了基于时间-空间域的联合图学习方法。具体来说,以最小化多通道含噪语音信号在空间图上的平滑度、参考通道信号在帧内图上的平滑度、空间图的稀疏度和帧内图的稀疏度之和为目标,在图的稀疏性和图的有效性约束条件下,构建空间图和帧内图的联合优化问题,并通过CVX工具求解此优化问题,得到多通道语音信号的空间图和帧内图。

(3)基于学习到的空间图和帧内图,构建多通道语音信号的时间-空间联合图,通过联合图傅里叶变换,将多通道语音图信号从图域变换到图频率域,并采用FBF方法对多通道信号进行波束形成,从而实现语音增强。

(4)仿真结果表明,相较于传统的FBF方法,所提出的JGL-FBF方法能有效提升增强语音的SNR和PESQ。

1 GSP理论的相关知识

本节简要介绍图信号处理的相关理论,包括图信号及其平滑度的概念、图傅里叶变换(Graph Fourier transform, GFT)及联合时间-顶点傅里叶变换(Joint time-vertex Fourier transform,JFT)。

1.1 图信号

在GSP中,通常用图$G=(\boldsymbol{V},\boldsymbol{E},\boldsymbol{W})$来表示对于一个含有$N$个节点和$M$个边的图信号,其中$V=\{v_1,v_2,\cdots,v_N\}$表示图结构中顶点的集合;$\boldsymbol{E}=\{e_{ij}\}$表示各顶点间是否连接,若$e_{ij}=1$,则表示顶点$v_i$和$v_j$间存在连接的边;$\boldsymbol{W}$表示每条边对应的权重,其元素$w_{ij}$反映由顶点$v_j$指向$v_i$联系强弱的程度。

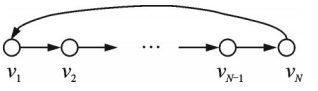

图1 有向周期循环图表示

Fig.1 Representation of the directed periodic cycle graph

以经典的有向周期循环图[2]为例,如图1所示。

图1中的图邻接矩阵是$N\times N$维的矩阵$\boldsymbol{A}$,元素$a_{ij}=1$表示节点i和节点j之间存在因果关系,由式(1)给出。

$$\boldsymbol{A}=\begin{bmatrix}0&0&0&\cdots&1\\1&0&0&&\vdots\\0&1&0&\ddots&0\\\vdots&&\ddots&\ddots&0\\0&\cdots&0&1&0\end{bmatrix} \tag{1}$$

基于图G,图信号$\boldsymbol{f}$的平滑度可以用图拉普拉斯的二次型来量化表示为[17]

$$\boldsymbol{f}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{f}=\frac{1}{2}\sum_{i,j}w_{ij}\big(f(i)-f(j)\big)^2 \tag{2}$$

式中:w_{ij}表示连接两个相邻顶点i和j的边上的权重,$f(i)$和$f(j)$分别为这两个顶点上的信号值,L为图拉普拉斯阵。假设权重为非负值,如果两个强连接的顶点(它们连接的边上有较大权重)具有相似值,则认为图信号$\boldsymbol{f}$是平滑的。式(2)表示的二次型值越小,图上的信号越平滑。

1.2 GFT

在图信号的描述中,有两种典型的权重矩阵表示形式。一种是描述有向图的邻接矩阵,另一种是描述无向图的拉普拉斯矩阵。对邻接矩阵或拉普拉斯矩阵进行特征分解,实现其对角化,将图信号从图域变换到图频域,进而分析信号在图频域的图频谱分布及相关特性。下面以邻接矩阵为例进行说明。

记$\boldsymbol{A}$是图信号$\boldsymbol{X}\in\mathbb{R}^{1\times N}$的邻接矩阵,对$\boldsymbol{A}$进行特征分解,得到

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1} \tag{3}$$

式中:$\boldsymbol{\Lambda}=\mathrm{diag}[\lambda_0,\lambda_1,\cdots,\lambda_{N-1}]$为$\boldsymbol{A}$的特征值矩阵,$\lambda_i(i=0,1,\cdots,N-1)$为信号的图频率,$\boldsymbol{U}$为$\boldsymbol{A}$的特征向量矩阵。用$\boldsymbol{U}^{-1}$作为图傅里叶矩阵$\boldsymbol{F}$[2],即

$$\boldsymbol{F}=\boldsymbol{U}^{-1} \tag{4}$$

对图信号$\boldsymbol{X}$进行GFT,得到图频域信号$\hat{\boldsymbol{X}}$为

$$\hat{\boldsymbol{X}}=\boldsymbol{F}\boldsymbol{X} \tag{5}$$

相应地,对$\hat{\boldsymbol{X}}$进行逆GFT(Inverse GFT,IGFT),可恢复图信号$\boldsymbol{X}$为

$$\boldsymbol{X}=\boldsymbol{F}^{-1}\hat{\boldsymbol{X}} \tag{6}$$

1.3 JFT

在GFT的基础上,将信号的时间维度也考虑在内,可采用JFT对时变图信号进行频率分析,主要思想是沿着时间域和顶点域进行谐波分析[24]。具体地,在顶点域上运用GFT和在时域上运用离散傅里叶变换(Discrete Fourier transform, DFT)来对信号分解变换,即

$$\tilde{\boldsymbol{X}}=\mathrm{JFT}\{\boldsymbol{X}\}=\boldsymbol{V}_{\mathrm{G}}^{-1}\boldsymbol{X}(\boldsymbol{V}_{\mathrm{T}})^{-1} \tag{7}$$

式中:$\boldsymbol{V}_{\mathrm{T}}$为DFT的基,$\boldsymbol{V}_{\mathrm{G}}$为GFT的基。

根据克罗内克积的性质$(\boldsymbol{V}_{\mathrm{T}}\otimes\boldsymbol{V}_{\mathrm{G}})x=\mathrm{vec}(\boldsymbol{V}_{\mathrm{G}}\boldsymbol{X}\boldsymbol{V}_{\mathrm{T}}^{\mathrm{T}})$,可得

$$\tilde{\boldsymbol{x}}=\mathrm{JFT}\{\boldsymbol{x}\}=(\boldsymbol{V}_{\mathrm{T}}\otimes\boldsymbol{V}_{\mathrm{G}})^{-1}\boldsymbol{x}=\boldsymbol{V}_{\mathrm{J}}^{-1}\boldsymbol{x} \tag{8}$$

式中:$\boldsymbol{x}=\mathrm{vec}(\boldsymbol{X})$为$\boldsymbol{X}$的矢量化形式,$\boldsymbol{V}_{\mathrm{J}}=\boldsymbol{V}_{\mathrm{T}}\otimes\boldsymbol{V}_{\mathrm{G}}$为JFT的基。相应地,对$\tilde{\boldsymbol{X}}$进行逆JFT(Inverse JFT,IJFT),可恢复图信号$\boldsymbol{X}$为

$$\boldsymbol{X}=\mathrm{IJFT}\{\tilde{\boldsymbol{X}}\}=\boldsymbol{V}_{\mathrm{G}}\tilde{\boldsymbol{X}}\boldsymbol{V}_{\mathrm{T}}^{\mathrm{T}} \tag{9}$$

值得提出的是,在上述JFT中,GFT和DFT运算的次序可交换,即对时间顶点信号进行GFT和DFT运算的次序不影响JFT结果。于是有

$$\mathrm{JFT}\{\boldsymbol{X}\}=\mathrm{GFT}\{\mathrm{DFT}\{\boldsymbol{X}\}\}=\mathrm{DFT}\{\mathrm{GFT}\{\boldsymbol{X}\}\} \tag{10}$$

2 基于联合图学习的多通道语音增强方法

本节提出一种基于联合图学习的多通道语音增强方法。首先,根据声场实际环境,设置一个参考麦克风,计算声源到其他麦克风的相对时延,进行时间延迟补偿,使得各通道信号在时间上对齐。然后,面向对齐后的多通道语音信号,设计联合图学习算法,构建三维语音图拓扑结构,再进行图频率域变换,最后通过波束形成得到增强后的语音。

2.1 基于联合图学习的三维语音图拓扑结构

针对多通道语音信号,研究构建三维语音图拓扑结构。具体地,基于麦克风阵列各阵元的相对关系如图2所示,构建表征通道间关系的空间图G_S;基于语音采样点间的结构关系,构建表征语音帧内各

样点关系的时间图G_T。联合G_S和G_T,构建三维语音图拓扑结构,如图3所示。

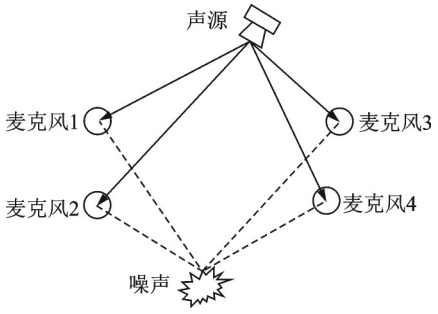

图2 麦克风阵列示意图

Fig.2 Microphone array diagram

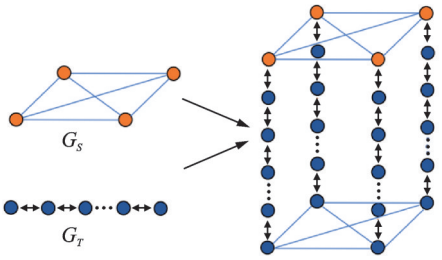

图3 三维语音图拓扑结构

Fig.3 Three dimensional speech graph topology

在空间图G_S中,把各个麦克风接收到的一帧语音信号视为一个顶点,对于一个由m个麦克风组成的阵列,麦克风底层拓扑结构可通过图$G_S=(V_S,E_S,W_S)$表示。其中V_S与每个麦克风相对应,E_S表示图上各顶点是否连接,W_S表示连接边对应的权重。在本文中仅考虑麦克风的空间相关性,故采用拉普拉斯矩阵L作为权重矩阵。通过构建麦克风空间图结构,除了关注接收信号本身,还同时考虑了各通道接收信号之间的联系。值得注意的是,在实际场景中,声源位置、环境噪声等因素可能在不断变化,若利用麦克风阵列的空间物理分布或其他来构建固定的空间图,并不能充分反映各通道接收信号间的关系随时间变化的特性,因此考虑采用图学习的方法,从多麦克风采集到的信号中学习此图拓扑结构。

在时间图G_T中,对于帧长为n的一帧语音信号$\boldsymbol{x}$,可通过图$G_T=(V_T,E_T,A)$映射到图域上成为语音图信号$\boldsymbol{x}_G$。V_T表示图上顶点的集合,与语音信号的采样点一一对应,同时,语音信号$\boldsymbol{X}$映射到图结构的顶点上成为语音图信号$\boldsymbol{x}_G$。E_T表示图结构中边的集合,A表示每条边对应的权重。考虑到语音信号的时序性,故采用邻接矩阵A作为权重矩阵。

基于上述所述,本文采用联合图学习的方法来同时构建L和A的联合优化问题。记$\boldsymbol{S}$是由不同通道同一帧信号组成的$m\times n$维矩阵,$\boldsymbol{Y}$是参考麦克风的一帧信号。根据以最小化多通道含噪语音信号在空间图上的平滑度、参考通道信号在语音帧内图上的平滑度构建图学习优化问题如下

$$\begin{cases}\min \alpha_1\operatorname{trace}\left(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{S}\right)+\alpha_2\operatorname{trace}\left(\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}^{\mathrm{T}}\right)+\beta_1\|\boldsymbol{L}\|_{\mathrm{F}}^2+\beta_2\|\boldsymbol{A}\|_{\mathrm{F}}^2\\ \text{s.t.}\quad \operatorname{trace}(\boldsymbol{L})=m\\ \quad\quad \operatorname{trace}(\boldsymbol{A})=n\\ \quad\quad \boldsymbol{L}*\boldsymbol{V}_1=0\\ \quad\quad \boldsymbol{L}_{ij}=\boldsymbol{L}_{ji}\\ \quad\quad \boldsymbol{A}*\boldsymbol{V}_2=\boldsymbol{V}_2\end{cases}\tag{11}$$

式中:α_1、α_2、β_1、β_2为正则化的参数,V_1为$m\times m$的全1阵,V_2为$n\times n$的全1阵。

在优化问题式(11)的目标函数中,$\operatorname{trace}\left(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{S}\right)$表示$\boldsymbol{S}$在空间图$\boldsymbol{L}$上的平滑度,$\operatorname{trace}\left(\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}^{\mathrm{T}}\right)$表示$\boldsymbol{Y}$在时间图$\boldsymbol{A}$上的平滑度,$\|\boldsymbol{L}\|_{\mathrm{F}}^2$和$\|\boldsymbol{A}\|_{\mathrm{F}}^2$分别表示$\boldsymbol{L}$的稀疏度和$\boldsymbol{A}$的稀疏度。值得提出的是,由于不同麦克风接收到的同一语音帧的帧内图结构通常基本一致,因此这里仅用参考麦克风接收到的信号$\boldsymbol{Y}$来学习语音帧内时间图$\boldsymbol{A}$。在优化问题式(11)的限制条件中,$\operatorname{trace}(\boldsymbol{L})=m$和$\operatorname{trace}(\boldsymbol{A})=n$是归一化因子,目的是为了避免平凡解和提高学习图的稀疏性,$\boldsymbol{L}*\boldsymbol{V}_1=0$和$\boldsymbol{L}_{ij}=\boldsymbol{L}_{ji}$分别保证$\boldsymbol{L}$是一个拉普拉斯阵和对称阵,$\boldsymbol{A}*\boldsymbol{V}_2=\boldsymbol{V}_2$用于$\boldsymbol{A}$的归一化。

式(11)为凸优化问题,可采用CVX工具进行求解,从而获得$\boldsymbol{L}$和$\boldsymbol{A}$。

2.2 图频域的多通道语音增强方法

将麦克风接收到的观测信号 $\boldsymbol{y}_i(t)(i=1,2,\cdots,m)$ 进行分帧加窗处理，按照帧号对不同通道的信号进行组合，构成 $m\times n$ 维的矩阵 $\boldsymbol{S}$。其中，m 表示麦克风的个数，n 为帧长，为表示方便，这里省略下标帧号。

对 $\boldsymbol{L}$ 和 $\boldsymbol{A}$ 分别进行特征分解，得到

$$\boldsymbol{L}=\boldsymbol{U}_S\boldsymbol{\Lambda}_S\boldsymbol{U}_S^{-1} \tag{12}$$

$$\boldsymbol{A}=\boldsymbol{U}_T\boldsymbol{\Lambda}_T\boldsymbol{U}_T^{-1} \tag{13}$$

结合 $\boldsymbol{U}_S$ 和 $\boldsymbol{U}_T$，对 $\boldsymbol{S}$ 进行联合图傅里叶变换(Joint graph Fourier transform, JGFT)，得到其图频谱 $\tilde{\boldsymbol{S}}$ 为

$$\tilde{\boldsymbol{S}}=\mathrm{JGFT}\{\boldsymbol{S}\}=(\boldsymbol{U}_S)^{-1}\times\boldsymbol{S}\times(\boldsymbol{U}_T)^{-1} \tag{14}$$

采用波束形成方法，在图频域对多通道语音图信号进行增强。具体地，将每个麦克风的图频谱乘上相应权值系数，然后相加求和，得到语音增强后的图频谱

$$\tilde{\boldsymbol{Z}}=\boldsymbol{w}^{\mathrm{H}}\tilde{\boldsymbol{S}} \tag{15}$$

式中 $\boldsymbol{w}$ 表示麦克风阵列波束形成的权重。

采用逆JGFT，可由增强语音图信号的图频谱恢复为增强的语音图信号

$$\boldsymbol{Z}=\mathrm{JGFT}^{-1}\left\{\tilde{\boldsymbol{Z}}\right\}=\boldsymbol{U}_S\times\tilde{\boldsymbol{Z}}\times\boldsymbol{U}_T \tag{16}$$

图4给出了基于联合图学习的多通道语音增强的系统框图。

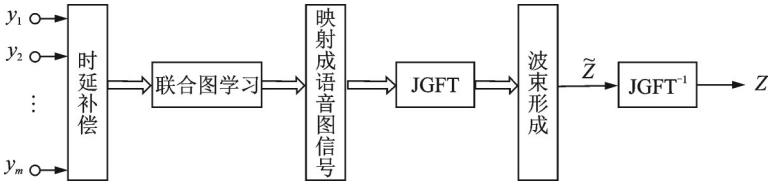

图 4 基于联合图学习的多通道语音增强的系统框图

Fig.4 Block diagram of multi-channel speech enhancement system based on joint graph learning

3 仿真与实验

本节仿真了所提出的JGL-FBF的语音增强性能，为验证本文提出的基于时间-空间域的联合图学习方法在麦克风阵列语音增强中的性能，选用传统的FBF方法作为对比方案。在传统的FBF方法中，直接对时域的多通道语音信号进行固定波束形成。同时，为了进一步验证约束条件的合理性，本文对单组约束条件进行了对比实验，只设置时间图学习的约束条件，不设置空间图学习的约束条件，即构建基于不完整约束条件的联合图学习问题(JGL under incomplete constraints，JGL-IC)。评价指标为增强语音的SNR和PESQ。

选取声学环境房间的大小为4 m×4 m×3 m，麦克风阵列为4个全向麦克风组成的均匀线性阵列，声源处在(2 m，3.5 m，1 m)位置处，混响时间为100 ms。为满足空间采样定理，阵元间距取12 cm，阵元分别处于(1.82 m，2 m，1 m)、(1.94 m，2 m，1 m)、(2.06 m，2 m，1 m)、(2.18 m，2 m，1 m)处。纯净语音信号取自TIMIT语音库，语音信号长度约为30 s，由10个说话人(5名男性和5名女性)组成，采样率为16 kHz。使用RIR生成器来产生多通道语音。散射噪声选用白噪声和高斯色噪声这两种噪声源。其中，高斯色噪声由高斯白噪声经过线性成形滤波器产生。在实验中，对含噪语音进行加窗分帧处理，所加窗为汉明窗，帧长为128样点，帧间重叠为50%。在FBF中，$w_1=w_2=w_3=w_4=\dfrac{1}{4}$。联合图学习的参数设置为：$\alpha_1=5,\alpha_2=3,\beta_1=0.9,\beta_2=0.8,m=4,n=128$。

表1给出了白噪声环境下所提JGL-FBF方法、传统FBF方法和JGL-IC方法的输出SNR。从表1可以看出,当输入SNR在－5、0、5和10 dB时,本文所提JGL-FBF方法的语音增强性能明显优于传统FBF方法和JGL-IC方法。当输入SNR在－10 dB时,本文所提JGL-FBF方法的输出SNR比FBF方法高。当输入SNR为10 dB时,JGL-IC方法的输出SNR甚至低于输入SNR。另外,从表1还可以看出,随着输入SNR的增加,3种方法的SNR提升均有所下降。

表1 白噪声环境下不同方法的输出SNR

Table 1 Output SNR of different methods in white noise environment

方法	输入信噪比/dB				
	－10	－5	0	5	10
MIC	－10 .001	－5.001	－0.004	4.986	9.955
FBF	－4.531	0.408	5.218	9.663	13.260
JGL-IC	－0.235	4.128	5.179	5.697	5.927
JGL-FBF	－3.225	4.240	8.283	11.144	13.569

从表2可以看出,本文提出的基于联合图学习的语音增强方法在PESQ得分上明显高于直接对输入信号进行波束形成,与JGL-IC方法的PESQ得分基本相等。当输入SNR在0和5 dB时,本文所提的JGL-FBF方法的PESQ得分提升比FBF方法超过了0.2。综合来看,在采用固定波束形成时,本文所提出的基于图学习的多通道语音增强方法在SNR和PESQ得分上均优于直接对输入信号进行波束形成方法和JGL-IC方法。

表2 白噪声环境下不同方法的PESQ

Table 2 PESQ of different methods in white noise environment

方法	输入信噪比/dB				
	－10	－5	0	5	10
MIC	1.169	1.286	1.494	1.816	2.176
FBF	1.243	1.496	1.848	2.212	2.583
JGL-IC	1.285	1.650	2.055	2.427	2.744
JGL-FBF	1.364	1.650	2.055	2.412	2.744

表3给出了高斯色噪声环境下所提JGL-FBF方法、传统FBF方法和JGL-IC方法的输出SNR。从表3可以看出,当输入SNR在0、5和10 dB时,本文所提JGL-FBF方法的语音增强性能明显优于传统FBF方法和JGL-IC方法。当输入SNR在－10和－5 dB时,本文所提JGL-FBF方法的输出SNR比FBF方法高。当输入SNR为10 dB时,JGL-IC方法的输出SNR甚至低于输入SNR。另外,从表3还可以看出,随着输入SNR的增加,3种方法的SNR提升均有所下降。

从表4可以看出,当输入SNR在－5、0、5和10 dB时,本文所提JGL-FBF方法的语音增强方法在PESQ得分上明显高于直接对输入信号进行波束形成,与JGL-IC方法的PESQ得分基本相等。当输入SNR在－10 dB时,3种方法下的PESQ得分基本接近。综合来看,在采用固定波束形成时,本文所提出的基于图学习的多通道语音增强方法在SNR和PESQ得分上均优于直接对输入信号进行波束形成方法和JGL-IC方法。

图5和图6分别仿真了不同噪声环境下SNR和PESQ得分与错位样点数N的关系曲线图,输入SNR为0 dB,混响时间为100 ms。从图5和图6可以看出,无论是在白噪声还是在高斯色噪声环境下,随着错位样点数N的增加,输出SNR和PESQ得分均明显下降。

表3 高斯色噪声环境下不同方法的输出SNR

Table 3 Output SNR of different methods in Gaussian color noise environment

方法	输入信噪比/dB				
	−10	−5	0	5	10
MIC	−10.000	−5.002	−0.004	4.985	9.958
FBF	−5.798	−0.846	4.008	8.576	12.439
JGL-IC	1.574	3.017	4.907	5.567	5.874
JGL-FBF	−1.421	3.017	7.126	10.600	13.032

表4 高斯色噪声环境下不同方法的PESQ

Table 4 PESQ of different methods in Gaussian color noise environment

方法	输入信噪比/dB				
	−10	−5	0	5	10
MIC	1.133	1.229	1.396	1.685	2.036
FBF	1.156	1.340	1.620	1.979	2.339
JGL-IC	1.150	1.447	1.836	2.220	2.568
JGL-FBF	1.151	1.448	1.836	2.220	2.567

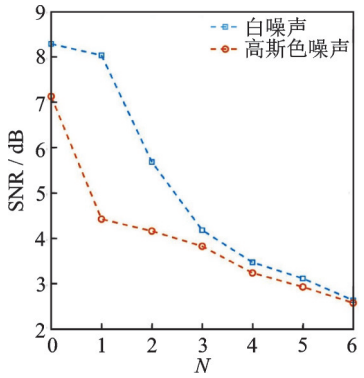

图5 不同噪声环境下输出SNR与错位样点数N的关系曲线图

Fig.5 Relationship curves of output SNR and the number of dislocation samples N in different noise environments

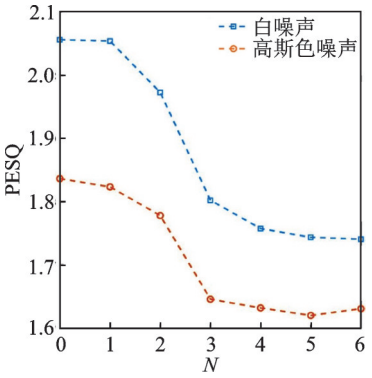

图6 不同噪声环境下PESQ与错位样点数N的关系曲线图

Fig.6 Relationship curves of PESQ and the number of dislocation samples N in different noise environments

本文研究了白噪声环境下所提JGL-FBF方法、传统FBF方法和JGL-IC方法的平均运行时间，输入SNR为0 dB，混响时间为100 ms。传统FBF方法的平均运行时间为0.015 s，JGL-FBF方法的平均运行时间为1 696.607 s，JGL-IC方法的平均运行时间为1 777.052 s。该结果由运行在内存为16 GB、CPU为Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz 2.59 GHz的主机上的MATLAB 2018b获得。可以看出，JGL-FBF的运行时间要比传统FBF方法长，其原因是在本文所提出的JGL-FBF方法中，语音信号每一帧都需要进行时间图和空间图的联合图学习，JGFT和JGFT^{-1}都会进行矩阵的高维度运算，这会导致JGL-FBF方法比单独的FBF方法花费更多的算力。随着快速JGFT和JGFT^{-1}的进一步研究，算法运行时间有望进一步降低。同时值得提出的是，相比静态图，联合图学习在揭示刻画动态图拓扑结构方面更具优势。

图7和图8分别给出了白噪声和高斯色噪声环境下,输入SNR为−5 dB时不同方法处理前后语音的时域波形图,混响时间为100 ms。从图7和图8可以看出,无论在白噪声环境还是高斯色噪声环境,本文提出的方法对噪声抑制效果明显更好。

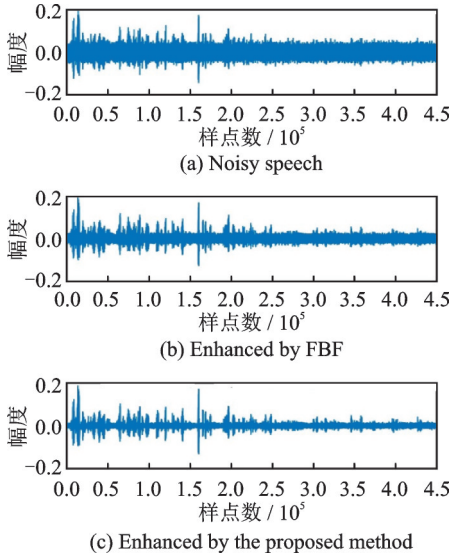

图7 白噪声环境下不同方法下增强语音的时域波形图

Fig.7 Time-domain waveforms of speech enhancement with different methods in white noise environment

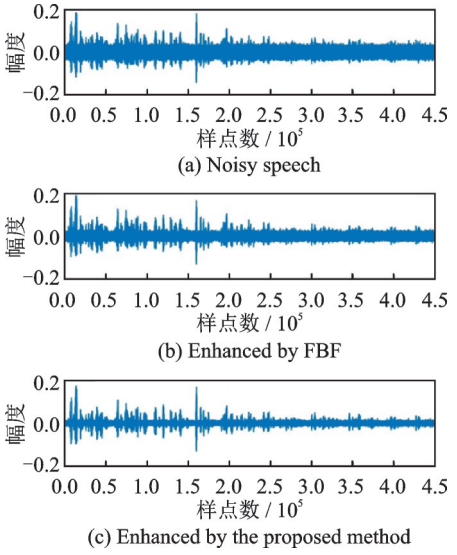

图8 高斯色噪声环境下不同方法下增强语音的时域波形图

Fig.8 Time-domain waveforms of speech enhancement with different methods in Gaussian color noise environment

4 结束语

本文提出了一种基于联合图学习的多通道语音增强方法。具体地,提出了一种联合时间-空间图学习方法,来表征多通道语音信号时间和空间上的内在关系。基于学习的空间图和帧内图,构建多通道语音信号的联合时间-空间图。通过JGFT,将多通道语音图信号变换到图频率域,得到包含语音帧内和通道间信息的预处理信号;然后利用FBF技术对其进行波束形成,得到语音增强后的信号。实验结果表明,所提出的基于联合图学习的多通道语音增强方法在SNR和PESQ得分上均有显著提升。另外,实验结果也表明,各路信号的时延补偿偏差会影响提出的基于联合图学习的多通道语音增强性能。

参考文献:

[1] VARMA R, CHEN S, KOVACEVIC J. Graph topology recovery for regular and irregular graphs[C]//Proceedings of 2017 IEEE 7th International Workshop on Computational Advances in Multi-sensor Adaptive Processing. Curacao, Netherlands: IEEE, 2017: 1-5.

[2] SANDRYHAILA A, MOURA J. Discrete signal processing on graphs: Frequency analysis[J]. IEEE Transactions on Signal Processing, 2014, 62(12): 3042-3054.

[3] ISUFI E, LOUKAS A, SIMONETTO A, et al. Autoregressive moving average graph filtering[J]. IEEE Transactions on Signal Processing, 2017, 65(2): 274-288.

[4] LOUKAS A, SIMONETTO A, LEUS G. Distributed autoregressive moving average graph filters[J]. IEEE Signal Processing Letters, 2015, 22(11): 1931-1935.

[5] YAGAN A C, OZGEN M T. Spectral graph based vertex-frequency wiener filtering for image and graph signal denoising[J]. IEEE Transactions on Signal and Information Processing over Networks, 2020, 6: 226-240.

[6] CHEN S, SANDRYHAILA A, KOVACEVIC J. Sampling theory for graph signals[C]//Proceedings of 2015 IEEE

International Conference on Acoustics, Speech and Signal Processing. Australian:IEEE, 2015: 3392-3396.
[7] 杨立山, 游康勇, 郭文彬. 基于扩散算子的带限图信号加权重建策略[J]. 电子与信息学报, 2017, 39(12): 2937-2944.
YANG Lishan, YOU Kangyong, GUO Wenbin. Weighted reconstruction strategy of band-limited graph signal based on diffusion operator[J]. Journal of Electronics & Information Technology, 2017, 39(12): 2937-2944.
[8] CHEN S, ELDAR Y C, ZHAO L. Graph unrolling networks: Interpretable neural networks for graph signal denoising[J]. IEEE Transactions on Signal Processing, 2021, 69: 3699-3713.
[9] EINIZADE A, SARDOUIE S H, SHAMSOLLAHI M B. Simultaneous graph learning and blind separation of graph signal sources[J]. IEEE Signal Processing Letters, 2021, 28: 1495-1499.
[10] 蒋俊正, 杨杰, 欧阳缮. 一种新的无线传感器网络中异常节点检测定位算法[J]. 电子与信息学报, 2018, 40(10): 2358-2364.
JIANG Junzheng, YANG Jie, OUYANG Shan. A new algorithm for abnormal node detection and location in wireless sensor networks[J]. Journal of Electronics & Information Technology, 2018, 40(10): 2358-2364.
[11] XU L, HUANG D, ZAIDI S F, et al. Graph Fourier transform based audio zero-watermarking[J]. IEEE Signal Processing Letters, 2021, 28: 1943-1947.
[12] PROUDLER I K, STANKOVIC V, WEISS S. Narrowband angle of arrival estimation exploiting graph topology and graph signals[C]//Proceedings of 2020 Sensor Signal Processing for Defence Conference. UK: SSPD, 2020: 1-5.
[13] YAN X, YANG Z, WANG T T, et al. An iterative graph spectral subtraction method for speech enhancement[J]. Speech Communication, 2020, 123: 35-42.
[14] WANG T T, GUO H Y, LYU B, et al. Speech signal processing on graphs: Graph topology, graph frequency analysis and denoising[J]. Chinese Journal of Electronics, 2020, 29(5): 138-148.
[15] 杨洋, 杨震. 基于图信号处理的MVDR波束形成多通道语音增强[J]. 南京邮电大学学报(自然科学版), 2021, 41(6): 35-40.
YANG Yang, YANG Zhen. MVDR beamforming multi-channel speech enhancement based on graph signal processing[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science), 2021, 41(6): 35-40.
[16] WITTEN D M, FRIEDMAN J H, SIMON N. New insights and faster computations for the graphical lasso[J]. Journal of Computational & Graphical Statistics, 2011, 20(4): 892-900.
[17] DONG X, THANOU D, FROSSARD P, et al. Learning Laplacian matrix in smooth graph signal representations[J]. IEEE Transactions on Signal Processing, 2016, 64(23): 6160-6173.
[18] EGILMEZ H E, PAVEZ E, ORTEGA A. Graph learning from data under laplacian and structural constraints[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(6): 825-841.
[19] RABBAT M G. Inferring sparse graphs from smooth signals with theoretical guarantees[C]//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. [S.l.]: IEEE, 2017: 6533-6537.
[20] CHEPURI S P, LIU S, LEUS G, et al. Learning sparse graphs under smoothness prior[C]//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. [S.l.]: IEEE, 2017: 6508-6512.
[21] SUBBAREDDY B, SIRIPURAM A, ZHANG J. Graph learning under spectral sparsity constraints[C]//Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. [S.l.]: IEEE, 2021: 5405-5409.
[22] KADAMBARI S K, CHEPURI S P. Product graph learning from multi-domain data with sparsity and rank constraints[J]. IEEE Transactions on Signal Processing, 2021, 69: 5665-5680.
[23] LI R, WANG J, XU W, et al. Graph laplacian matrix learning from smooth time-vertex signal[J]. China Communications, 2021, 18(3): 187-204.
[24] GRASSI F, LOUKAS A, PERRAUDIN N, et al. A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs[J]. IEEE Transactions on Signal Processing, 2018, 66(3): 817-829.

作者简介:

张鹏程(1998-),男,硕士研究生,研究方向:图信号处理、语音信号处理等,E-mail: 1220013925@njupt.edu.cn。

郭海燕(1983-),女,副教授,硕士生导师,研究方向:语音信号处理、B5G/6G无线传输等。

王婷婷(1992-),女,博士研究生,研究方向:图信号处理、语音信号处理等。

杨震(1961-),**通信作者**,男,教授,博士生导师,研究方向:语音处理与现代语音通信、无线通信中的通信与信号处理技术等,E-mail: yangz@njupt.edu.cn。

(编辑:夏道家)