

结合脉冲响应重塑和期望最大化的盲信号分离

解 元^{1,2}, 张 旭¹, 邹 涛^{1,3}, 马 鸽^{1,4}, 余锦视¹, 孙为军^{5,6}

(1. 广州大学机械与电气工程学院, 广东广州 510006; 2. 智能检测与制造物联教育部重点实验室, 广东广州 510006; 3. 粤港澳复杂制造多尺度信息融合与协同优化控制重点实验室, 广东广州 510006; 4. 广州市制造过程综合自动化重点实验室, 广东广州 510006; 5. 广东省物联网信息技术重点实验室, 广东广州 510006; 6. 物联网智能信息处理与系统集成教育部重点实验室, 广东广州 510006)

摘 要: 多通道欠定卷积语音混合信号的分离问题是盲信号分离领域的难点. 由于混合信号中常伴随声学回声和混响, 真实的源信号很难完全被清晰地分离出来. 传统的盲信号分离算法多数适用于低混响, 而在高混响场景下, 算法的分离性能极速下降甚至是失效的. 本文针对具有声学回声和混响环境下的多通道欠定卷积语音混合信号的分离问题, 提出一种结合脉冲响应重塑和期望最大化的盲信号分离算法, 该算法在低混响和高混响下都表现出很好的分离性能. 首先, 利用基于无穷范数和 p -范数的脉冲响应重塑技术设计预滤波器消除可听回声, 完成对混合信号的重塑, 提高混合信号的质量. 然后, 对重塑后的混合信号利用分层聚类方法估计混合矩阵, 基于期望最大化算法框架, 设计新的模型参数实时更新规则, 通过结合脉冲响应重塑和期望最大化重构源信号. 实验结果表明, 所提算法可以有效地分离不同混响环境下带声学回声的欠定卷积混合信号, 其分离性能优越, 同时对噪声具有很好的鲁棒性.

关键词: 盲源分离; 脉冲响应重塑; 多通道卷积混合; 期望最大化

基金项目: 广州市基础与应用基础研究项目 (No.SL2022A04J00289); 广东省基础与应用基础研究基金 (No.2023A1515011311); 国家自然科学基金 (No.62003095, No.52171331); 广州市市校联合实验室项目 (No.2023A03J0120)

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112(2023)11-3343-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230272

Blind Signal Separation Combining Impulse Response Remodeling and Expectation Maximization

XIE Yuan^{1,2}, ZHANG Xu¹, ZOU Tao^{1,3}, MA Ge^{1,4}, YU Jin-shi¹, SUN Wei-jun^{5,6}

(1. School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, Guangdong 510006, China;

2. Key Laboratory of Intelligent Detection and the Internet of Things in Manufacturing, Ministry of Education, Guangzhou, Guangdong 510006, China;

3. Guangdong-Hong Kong-Macao Key Laboratory of Multi-Scale Information Fusion and Collaborative Optimization Control of Complex Manufacturing Process, Guangzhou, Guangdong 510006, China;

4. Guangzhou Key Laboratory of Manufacturing Process Comprehensive Automation, Guangzhou, Guangdong 510006, China;

5. Guangdong Provincial Key Laboratory of Information Technology of Internet of Things, Guangzhou, Guangdong 510006, China;

6. Key Laboratory of Intelligent Information Processing and System Integration of Internet of Things, Ministry of Education, Guangzhou, Guangdong 510006, China)

Abstract: The separation of multichannel underdetermined convolutive speech mixing signals is a difficult problem in the field of blind signal separation. Due to the acoustic echo and reverberation in the mixing signals, it is difficult to completely and clearly separate the real source signals. Most traditional blind signal separation algorithms are suitable for low reverberation, while in high reverberation scenarios, the separation performance of the algorithm rapidly degrades or even fails. To separate multichannel convolutive mixing signals with acoustic echo and reverberation, a blind signal separation algorithm is proposed combining impulse response remodeling and expectation maximization, which exhibits good separation performance in both low and high reverberation environments. Firstly, a pre-filter is designed using impulse response remodeling technology based on infinite and p -norm to eliminate audible echoes, remodeling the room impulse response and improving the quality of the mixing signals. Then, a hierarchical clustering method is used to estimate the mixing ma-

trix for the remodeled mixing signals, the new model parameters real-time update rules are designed based on expectation maximization algorithm framework, and the source signals are reconstructed combining the impulse response remodeling and expectation maximization. Experimental results show that the proposed algorithm can effectively separate the speech mixing signals with acoustic echoes in different reverberation environments, owning superior separation performance and good robustness to noise.

Key words: blind signal separation; impulse response remodeling; multichannel convolutive mixtures; expectation maximization

Foundation Item(s): Guangzhou Basic and Applied Basic Research Project (No.SL2022A04J00289); Guangdong Provincial Basic and Applied Basic Research Fund (No.2023A1515011311); National Natural Science Foundation of China (No.62003095, No.52171331); Guangzhou City School Joint Laboratory Project (No.2023A03J0120)

1 引言

盲信号分离的目的是从接收的混合信号中将源信号独立分离出来,其中“盲”的定义是源信号和混叠通道的信息未知,其最早起源于鸡尾酒会问题^[1,2]. Jutten和Herault发表于1991年的论文中提出了一种基于神经网络的盲分离方法^[3],自此,盲源分离问题得到广泛关注.随着盲源分离相关理论的发展,该项技术已经在机械故障诊断、文本数据挖掘、图像处理、语音信号处理和生物信息等领域中得到广泛应用^[4-7].本文关注的是语音信号处理领域中,针对具有声学回声和混响环境下的多通道卷积语音混合信号的盲分离问题.

根据应用场景中不同的麦克风和声源数目,可将盲信号分离问题分为三类:第一类为正定情况,即麦克风数目和声源数目相同;第二类为超定情况,即麦克风数目大于声源数目;第三类为欠定情况,即麦克风数小于声源数目.针对正定和超定情况,独立成分分析(Independent Component Analysis, ICA)是最经典的方法,其通过使输出信号之间的独立性最大化完成对源信号的提取^[8-10].针对欠定混叠,可以采用稀疏成分分析法(Sparse Component Analysis, SCA),该方法要求源信号满足一定的稀疏性,利用稀疏表示的算法处理盲信号分离问题^[11,12].

传统的基于ICA和SCA盲源分离方法都是利用短时傅里叶变换(Short Time Fourier Transform, STFT),在线性窄带近似假设下进行的,要求声源对麦克风的脉冲响应长度远远短于STFT窗长度.该条件限制了算法的实用性,特别是在高混响条件下,脉冲响应长度可能超过STFT窗长度,导致信号分离性能急剧下降. Duong等^[13]提出一种满秩空间协方差模型(Full-rank spatial covariance model, Full-rank)算法, Feng等^[14]设计了一种卷积窄带近似的稀疏方法(Convolutive Proximal Alternating Linearized Minimization, C-PALM), Guo等^[15]提出一种基于掩蔽波束形成的欠定盲源分离方法.但这些方法在高混响环境下仍然很难适用,并且在分离混合信号的每个频率段都会出现任意尺度和排序,导致了

尺度和排序歧义性问题.为了解决尺度和排序歧义性问题,基于平行因子分析^[16]和波达方向(Direction Of Arrival, DOA)技术^[17-19]是比较经典的解决方法.近期也出现了一种基于张量分解模型,通过使用耦合频率箱来消除了排列歧义性^[20-22].然而,以上算法并未彻底解决排序歧义性问题,从而导致源信号的分离性能下降.

非负矩阵分解(Non-negative Matrix Factorization, NMF)是将一个非负的矩阵分解为左右两个低秩的非负矩阵的乘积^[23,24],基于NMF技术的盲源分离算法已经被广泛应用于语音信号处理^[25-28].该技术的优势是可以避免排序歧义性问题,并结合期望最大化^[29](Expectation-Maximization, EM)算法,可提高算法的收敛效率. Xie等^[30]提出了一种多通道欠定盲源分离(Multi-channel Underdetermined Blind Source Separation, MUBSS)算法,通过改进传统的EM算法,实现了运算更快速的收敛,但当接收的观测信号质量较差时,EM算法的分离性能仍不够理想,有待进一步提升.

为了解决具有声学回声和混响环境下的多通道卷积语音混合信号的盲源分离问题,本文先对混叠系统中的房间脉冲响应进行重塑^[31],然后在EM算法的框架下,设计一种结合脉冲响应重塑和期望最大化的盲信号分离算法,命名为IRR-EM(Impulse Response Remodeling-Expectation Maximization)算法.该算法先对接收到的混合信号进行预滤波处理,抑制了房间脉冲响应的早期混响部分,增大了信号混响能量比,消除可听回声,提高了接收信号的质量.预滤波过程完成后建立时频域混合信号模型,设计新的EM算法对混合信号进行盲分离,实现源信号的重构.仿真结果表明,所提的IRR-EM算法与流行的EM算法相比,源信号的分离性能得到了明显的提升,其在低混响和高混响环境下都表现出较好的分离性能,并且该算法具有很好的鲁棒性.

2 模型与算法流程

2.1 卷积混合信号模型

在实际的房间内进行声学信号的接收过程中,混

合信号中常常伴随一定的回声和混响. 因此,对混合信号建模为多通道卷积混合模型:

$$\mathbf{x}(t) = \sum_{u=1}^U \mathbf{A}_u \mathbf{s}(t) + \mathbf{b}(t) \quad (1)$$

$$\mathbf{A}_u = \begin{pmatrix} a_{11}^u \sigma(t - \tau_{11}^u) & \cdots & \cdots & a_{1j}^u \sigma(t - \tau_{1j}^u) \\ a_{21}^u \sigma(t - \tau_{21}^u) & \ddots & \cdots & a_{2j}^u \sigma(t - \tau_{2j}^u) \\ \vdots & & a_{lm}^u \sigma(t - \tau_{lm}^u) & \vdots \\ a_{i1}^u \sigma(t - \tau_{i1}^u) & \cdots & \cdots & a_{ij}^u \sigma(t - \tau_{ij}^u) \end{pmatrix} \quad (2)$$

其中, $\mathbf{x}(t)=[x_1(t), \dots, x_j(t)]^T$ 代表麦克风接收到的混合信号; $\mathbf{s}(t)=[s_1(t), \dots, s_l(t)]^T$ 代表源信号, l 代表信号源的数目; $\mathbf{b}(t)=[b_1(t), \dots, b_j(t)]^T$ 是噪声, J 代表麦克风的数目; U 代表空间信号传播路径数; \mathbf{A}_u 表示第 u 个传播路径下的混合矩阵; $\sigma(\cdot)$ 表示脉冲响应; a_{lm}^u 和 τ_{lm}^u 分别为振荡和时间延迟.

2.2 算法流程概述

本文提出一种结合脉冲响应重塑和期望最大化的盲信号分离算法,如图 1 所示. 第一阶段利用 IRR 技术对接收到的室内混响信号进行脉冲响应重塑,滤除信号中存在的回声和混响,作用是降低高混响影响,提高信号质量. 第二阶段是源信号的分离阶段,包括混合矩阵估计、构建非负矩阵分解源信号模型、设计模型参数实时更新规则,以及在新规则下重构频域源信号,进而提出结合脉冲响应重塑和期望最大化的盲信号分离算法(IRR-EM),实现源信号的分离.

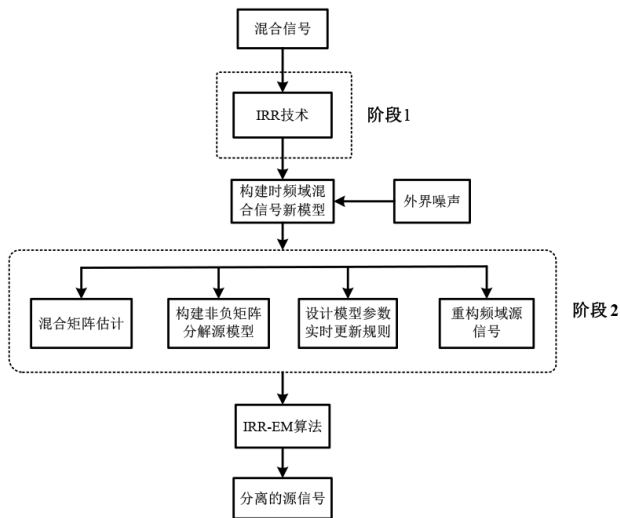


图 1 IRR-EM 算法流程图

3 算法

3.1 IRR 技术

应用 IRR 技术的目的是设计一个预滤波器,使全局脉冲响应 $\mathbf{g}(n)$ 不仅衰减得比房间的脉冲响应 $\mathbf{c}(n)$ 快,而且在满足声学条件下,可消除可听到的回声,提高信

号的懂度和质量. 为了设计一个优化的预滤波器,实现具有快速而单调的衰减特性的全局脉冲响应 $\mathbf{g}(n)$,将无穷和 p -范数准则与人类听觉系统的特性结合起来,完成对房间脉冲响应的重塑过程,提高接收信号的质量.

设计预滤波器 \mathbf{h} ,使其满足:

$$\mathbf{g}(n) = \mathbf{A}_u \mathbf{h} \quad (3)$$

其中, $\mathbf{g}(n)$ 是长度为 L_g 的全局脉冲响应, \mathbf{h} 是长度为 L_h 的预滤波器的脉冲响应.

3.1.1 窗函数设计

系统的脉冲响应及其频率响应的能量衰减特性依赖所选择的窗函数,因此窗函数的设计在整个重塑过程中起着重要作用. 全局脉冲响应 $\mathbf{g}(n)$ 以某种方式进行衰减,以至于没有可听到的回声,这意味着混响应该通过人类听觉系统的前向掩蔽效应被主声音掩盖. 人类听觉系统在真实声学环境中的前向掩蔽效应取决于所考虑的信号和房间脉冲响应所描述的房间特征^[32].

为了控制全局脉冲响应的衰减特性,本文使用两个窗口函数 \mathbf{w}_d 和 \mathbf{w}_u ,从全局脉冲响应中得到一个期望的部分 $\mathbf{g}_d = \mathbf{w}_d(n)\mathbf{g}(n)$ 和一个不期望的部分 $\mathbf{g}_u = \mathbf{w}_u(n)\mathbf{g}(n)$. 将两个窗函数定义如下:

$$\mathbf{w}_u = [0_1, 0_2, \dots, 0_{N_1+N_2}, \mathbf{w}_0^T, 0_{N_3}]^T \quad (4)$$

$$\mathbf{w}_d = [0_1, \dots, 0_{N_1}, 1_1, \dots, 1_{N_2}, 0_1, \dots, 0_{N_3}]^T \quad (5)$$

其中, $N_1 = t_0 f_s$, $N_2 = 0.004 f_s$ 和 $N_3 = L_g - N_1 - N_2$, f_s 为采样频率, t_0 表示声源的有效时间.

窗函数 \mathbf{w}_0 定义如下^[31]:

$$\mathbf{w}_0(n) = 10^{\frac{3[\log_{10}(\eta(N_1+N_2))]}{\log_{10}[N_0(N_1+N_2)]} + \frac{1}{2}} \quad (6)$$

其中, $N_0 = (0.2 + t_0) f_s$, 时间索引值 n 从 $N_1 + N_2 + 1$ 取值到 $L_g - 1$, $\mathbf{w}_0(n)$ 代表人类听觉系统的折中掩蔽极限. 根据对窗函数设计可以使得系统的脉冲响应衰减到人类听觉系统的折中掩蔽极限以下,从而消除可听回声.

3.1.2 基于无穷和 p -范数的 IRR 技术

为了对房间脉冲响应进行重塑,可以根据不同的准则以不同的方式设计所需的滤波器. 由于无穷范数在计算上具有高效率的优点, p -范数具有迭代收敛快速的特点,于是本文对全局脉冲响应期望的部分采用无穷范数优化,对全局脉冲响应不期望的部分采用 p -范数优化,给出了无穷和 p -范数准则优点的折中. 该设计使滤波器的设计误差均匀分布在房间脉冲响应重塑的时间范围内,同时与人类听觉系统的声学特性相结合,以此定义出优化准则. 利用该优化准则,不会产生后期

扩散混响,并且能够精确地控制声学回波。

本文对预滤波器的优化如下:

$$\text{MIN}_h : f(\mathbf{h}) = \ln \left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})} \right) \quad (7)$$

其中, $\ln(\cdot)$ 是自然对数,

$$f_d(\mathbf{h}) = \|\mathbf{g}_d\|_\infty = \text{Max}[\|\mathbf{g}_d\|] \quad (8)$$

$$f_u(\mathbf{h}) = \|\mathbf{g}_u\|_{p_u} = \left(\sum_{k=0}^{L_g-1} |g_u(k)|^{p_u} \right)^{\frac{1}{p_u}} \quad (9)$$

选择 $p_u = 10$. 利用梯度下降法获得更新:

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu \nabla_{\mathbf{h}} f(\mathbf{h}^l) \quad (10)$$

其中, $\mathbf{h}^l = [0.01, 0, \dots, 0]^T$,

$$\nabla_{\mathbf{h}} f(\mathbf{h}) = \frac{1}{f_u(\mathbf{h})} \nabla_{\mathbf{h}} f_u(\mathbf{h}) - \frac{1}{f_d(\mathbf{h})} \nabla_{\mathbf{h}} f_d(\mathbf{h}) \quad (11)$$

且

$$\begin{aligned} \nabla_{\mathbf{h}} f_u(\mathbf{h}) &= \left(\sum_{k=0}^{L_g-1} |g_u(k)|^{p_u} \right)^{\frac{1}{p_u}-1} \\ &\quad \times \text{diag}(\text{sign}(\mathbf{g}_u)) \text{diag}(\mathbf{w}_u) \mathbf{A}_u^T \mathbf{g}_u^{p_u-1} \\ \nabla_{\mathbf{h}} f_d(\mathbf{h}) &= \text{sign}(\mathbf{g}_d(I_d)) \mathbf{w}_d(I_d) \mathbf{A}_{I_d}^T \end{aligned} \quad (12)$$

其中, $\text{diag}(\cdot)$ 是对角化函数, $\text{sign}(\cdot)$ 是符号函数, I_d 是 $|\mathbf{g}_d(n)|$ 最大值所处的位置, \mathbf{A}_{I_d} 代表矩阵 \mathbf{A}_u 的第 I_d 行. 把式(11)、式(12)、式(13)带入式(10)中,并令 $\phi f_u = \sum_{k=0}^{L_g-1} |g_u(k)|^{p_u}$, $\mathbf{a}_u = \text{diag}(\text{sign}(\mathbf{g}_u)) \text{diag}(\mathbf{w}_u) \mathbf{A}_u^T \mathbf{g}_u^{p_u-1}$, 得到基于无穷范数和 p -范数准则的梯度下降更新:

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu \left(\frac{1}{\phi f_u} \mathbf{a}_u - \frac{1}{g(I_d)} \mathbf{A}_{I_d}^T \right) \quad (14)$$

通过对窗函数和预滤波器的优化设计,严格地控制了系统中脉冲响应的衰减特性和环境中的混响和回声,完成了对房间脉冲响应的重塑过程,由此得到了可懂度和质量更高的混叠信号 $\mathbf{x}(t)$.

3.2 盲源分离算法设计

利用STFT和瞬时线性窄带近似,将式(1)改写为

$$\mathbf{X}(f, n) = \mathbf{A}_f \mathbf{S}(f, n) + \mathbf{b}(f, n) \quad (15)$$

其中, $\mathbf{X}(f, n)$, $\mathbf{S}(f, n)$ 和 $\mathbf{b}(f, n)$ 分别表示 $\mathbf{x}(t)$, $\mathbf{s}(t)$ 和 $\mathbf{b}(t)$ 的短时傅里叶变换; \mathbf{A}_f 是 \mathbf{A}_u 经过重塑以后的短时傅里叶变换在频域下的混合矩阵; $f = 1, 2, \dots, F$ 和 $n = 1, 2, \dots, N$ 代表频点和时间帧. 因此,将室内混响环境下时域上的卷积混合转换为频域上各个频点的瞬时线性混合. 该时频域混合信号模型的优势在于经过上述IRR技术消除可听回声,降低高混响环境的影响,在此基础上设计盲源分离算法,可以有效地提高算法的分离性能.

为了得到更好的分离源信号,基于期望最大化合

法框架,设计一种新的盲源分离算法. 首先,通过分层聚类得到混合矩阵 \mathbf{A}_f . 然后,构建非负矩阵分解源信号模型;在期望最大化合算法框架下,设计新的模型参数更新规则,进而重构源信号.

3.2.1 混合矩阵估计

对混合信号进行STFT系数的向量归一化:

$$\bar{\mathbf{X}}(f, n) \leftarrow \frac{\mathbf{X}(f, n)}{\|\mathbf{X}(f, n)\|_2} e^{-i \cdot \arg(X_i(f, n))} \quad (16)$$

其中, $\|\cdot\|_2$ 代表欧氏范数, $\arg(\cdot)$ 表示复数的辐角. 假设在每个频率下,所有源信号的声音都来自不相交的空间区域,并且在大多数时频点处,只有一个源信号占主导地位^[33]. 然后,将混合信号的STFT系数向量在第 i 个源占主导地位的时间帧中聚集在相关联的混合向量的方向周围. 为了得到这些聚类,然后,定义两个聚类 C_1 和 C_2 之间的距离:

$$d(C_1, C_2) = \frac{1}{|C_1| \cdot |C_2|} \sum_{\bar{\mathbf{x}}_{c_1} \in C_1} \sum_{\bar{\mathbf{x}}_{c_2} \in C_2} \|\bar{\mathbf{x}}_{c_1} - \bar{\mathbf{x}}_{c_2}\|_2 \quad (17)$$

其中, $\bar{\mathbf{x}}_{c_1}$ 和 $\bar{\mathbf{x}}_{c_2}$ 表示混合STFT系数的归一化向量,每个归一化向量视为包含单个项的聚类,计算出每对簇之间的距离,并合并具有最小距离的两个簇,得到聚类 C_1 和 C_2 . 选择集群样本数量等于源信号数目的簇,计算出混合向量:

$$\mathbf{a}_{i,j}(f) = \frac{1}{|C_j|} \sum_{\bar{\mathbf{x}}(f,n) \in C_j} \mathbf{X}(f, n) e^{-i \cdot \arg(X_i(f, n))} \quad (18)$$

其中, $|C_j|$ 表示在频率 f 处簇 C_j 的样本总数,从而得出混合矩阵 \mathbf{A}_f . 然后,基于DOA的方法来优化排序歧义性问题.

3.2.2 源信号模型

源信号的功率谱满足非负矩阵分解,即

$$|\mathbf{S}_i|^2 \approx \mathbf{Z}_i \mathbf{H}_i \quad (19)$$

其中, $|\mathbf{S}_i|^2 \in \mathbf{R}^{F \times N}$ 代表第 i 个源信号的STFT矩阵, $\mathbf{Z}_i \in \mathbf{R}^{F \times K}$ 是非负的字典矩阵, $\mathbf{H}_i \in \mathbf{R}^{K \times N}$ 是非负的活跃矩阵. 假设在每个时频点上的源信号 $\mathbf{S}_i(f, n)$ 可以表示为多个成分 $c_k(f, n)$ 之和,满足:

$$\mathbf{S}_i(f, n) = \sum_{k=1}^K c_k(f, n) \quad (20)$$

且

$$c_k(f, n) \sim \mathbf{N}_c(0, z(f, n) h(f, n)) \quad (21)$$

因此,源信号 $\mathbf{S}_i(f, n)$ 满足:

$$\mathbf{S}_i(f, n) \sim \mathbf{N}_c \left(0, \sum_{k=1}^K z(f, k) h(k, n) \right) \quad (22)$$

其中, $z(f, k)$, $h(k, n)$ 分别代表非负矩阵 $\mathbf{Z}_i, \mathbf{H}_i$ 中的元素.

3.2.3 模型参数更新规则设计

为了分离源信号并获得更好的分离性能,可以通过对模型参数 $\theta = \{A(f, n), z(f, n), h(f, n)\}$ 进行实时更新,以此获得高精度的源信号分离,使算法的分离性能更加优越.

计算 $X(f, n)$ 的协方差矩阵,表示为

$$\begin{aligned} \sum_x(f, n) &= \mathbf{E}[X(f, n)X^H(f, n)] \\ &= A_f \sum_s(f, n) A_f^H + \sum_b(f, n) \end{aligned} \quad (23)$$

其中, $(\cdot)^H$ 表示矩阵的共轭转置, $\sum_s(f, n) = \text{diag}\left(\left[\sum_{k=1}^K z(f, n)h(f, n)\right]_i\right)$, $\sum_b(f, n)$ 表示噪声 $b(f, n)$ 的协方差矩阵. 定义目标函数:

$$C(\theta) = \sum_{f, n} \text{tr}\left(\frac{X(f, n)X^H(f, n)}{\sum_x(f, n)}\right) + \log_{10} \left| \sum_x(f, n) \right| \quad (24)$$

其中, $\text{tr}(\cdot)$ 表示矩阵的迹, 设计维纳滤波器:

$$W_s(f, n) = \frac{\sum_s(f, n) A_f^H}{\sum_x(f, n)} \quad (25)$$

$$W_c(f, n) = \frac{\sum_c(f, n) A_f^H}{\sum_x(f, n)} \quad (26)$$

其中, $\sum_c(f, n) = \text{diag}\left(\left[z(f, n)h(f, n)\right]_k\right)$, 则

$$\hat{S}(f, n) = W_s(f, n)X(f, n) \quad (27)$$

$$\hat{c}(f, n) = W_c(f, n)X(f, n) \quad (28)$$

定义

$$R_{ss}(f, n) = \frac{1}{N} \sum_{n=1}^N S(f, n)S^H(f, n) \quad (29)$$

$$R_{xs}(f, n) = \frac{1}{N} \sum_{n=1}^N X(f, n)S^H(f, n) \quad (30)$$

$$R_{xx}(f, n) = \frac{1}{N} \sum_{n=1}^N X(f, n)X^H(f, n) \quad (31)$$

$$v_k(f, n) = |c_k(f, n)|^2 \quad (32)$$

然后,设计EM算法在当前参数估计的基础上有条件地计算自然统计的期望值(E步骤),以及使用更新的自然统计重新估计模型参数(M步骤),具体如下:

E-step:

$$\begin{aligned} \hat{R}_{ss}(f, n) &= \frac{1}{N} \sum_{n=1}^N \hat{S}(f, n)\hat{S}^H(f, n) \\ &\quad + (I - W_s(f, n)A_f) \sum_s(f, n) \end{aligned} \quad (33)$$

$$\hat{R}_{xs}(f, n) = \frac{1}{N} \sum_{n=1}^N X(f, n)\hat{S}^H(f, n) \quad (34)$$

$$\hat{R}_{xx}(f, n) = \frac{1}{N} \sum_{n=1}^N X(f, n)X^H(f, n) \quad (35)$$

$$\hat{v}_k(f, n) = \left[\hat{c}(f, n)\hat{c}^H(f, n) + (I - W_c(f, n)A_f) \sum_c(f, n) \right]_{k, k} \quad (36)$$

其中, I 表示单位矩阵.

M-step:

$$\hat{A}_f = \frac{\hat{R}_{xs}(f)}{\hat{R}_{ss}(f)} \quad (37)$$

$$z(f, k) = \frac{1}{N} \sum_{n=1}^N \frac{\hat{v}_k(f, n)}{h(f, n)} \quad (38)$$

$$h(k, n) = \frac{1}{F} \sum_{f=1}^F \frac{\hat{v}_k(f, n)}{z(f, k)} \quad (39)$$

通过以上更新规则,对模型参数进行实时更新学习.

3.2.4 源信号重构

利用维纳滤波器获得频域中的估计源信号:

$$\begin{aligned} \hat{S}(f, n) &= \frac{\sum_s(f, n) \hat{A}_f^H}{\hat{A}_f \sum_s(f, n) \hat{A}_f^H} X(f, n) \\ &= \frac{\text{diag}\left(\left[\sum_{k=1}^K z(f, n)h(f, n)\right]_i\right) \hat{A}_f^H}{\hat{A}_f \text{diag}\left(\left[\sum_{k=1}^K z(f, n)h(f, n)\right]_i\right) \hat{A}_f^H} X(f, n) \end{aligned} \quad (40)$$

最后,利用STFT的逆运算把频域中估计的源信号变换到时域,实现源信号的分离.由于本文重点解决具有声学回声和混响环境下的多通道卷积语音混合信号的盲源分离问题,先对房间脉冲响应进行重塑,在此基础上设计新的期望最大化算法,进而提出一种结合脉冲响应重塑技术和期望最大化的盲源分离算法,所以把所设计的算法命名为IRR-EM算法.

4 实验

为了验证所提的IRR-EM算法的有效性,分别对正定和欠定情况下的中英文语音信号进行盲源分离实验.首先,对实验参数进行设置,选择算法的分离性能评估准则.然后,选择从国际上公开的SiSEC'08数据集(<http://www.sisec.wiki.irisa.fr>)和国内公开的AISHELL数据集(<http://www.aishelltech.com/kysjcp>)中分别获得英文语音信号和中文语音信号进行盲信号分离实验.最后,对算法的鲁棒性进行实验分析.所有实验都是在Windows 10操作系统,2.20 GHz、8.00 GB内存,AMD Ryzen 5 3550H处理器,Matlab R2018b环境下对程序进行的编码.

4.1 实验参数

实验中共同的实验参数如表1所示.

表 1 实验参数设置

参数	参数值
实验空间大小	4.5 m × 3.5 m × 2.5 m
源信号的数目	3, 4
麦克风的数目	2
混响时间(RT60)	100~700 ms
采样频率(Fs)	16 kHz
麦克风间距	5 cm
窗函数	Hanning window
短时傅里叶变换窗长度	2 048(128 ms)
房间脉冲响应长度	RT60×Fs

4.2 分离性能评估准则

为了评估所提算法的分离性能,本文使用了专门用于评价盲源分离算法性能的工具箱 `bss_eval`. 通过工具箱计算出信号源失真比(Source-to-Distortion Ratio, SDR)、源干扰比(Source-to-Interference Ratio, SIR)和源伪影比(Signal-to-Artifact Ratio, SAR)对算法的分离性能进行评估^[33]. 分离出的源信号可以表示为

$$\hat{s}_i = s_{\text{target}} + e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}} \quad (41)$$

其中, $1 \leq i \leq I$, s_{target} 为目标信号, e_{interf} 为干扰信号, e_{noise} 为噪音信号, e_{artif} 为伪源信号. 评价准则 SDR, SIR, SAR 定义为

$$\text{SDR} = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}}\|^2} \quad (42)$$

$$\text{SIR} = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}}\|^2} \quad (43)$$

$$\text{SAR} = 10 \log_{10} \frac{\|s_{\text{target}} + e_{\text{interf}} + e_{\text{noise}}\|^2}{\|e_{\text{artif}}\|^2} \quad (44)$$

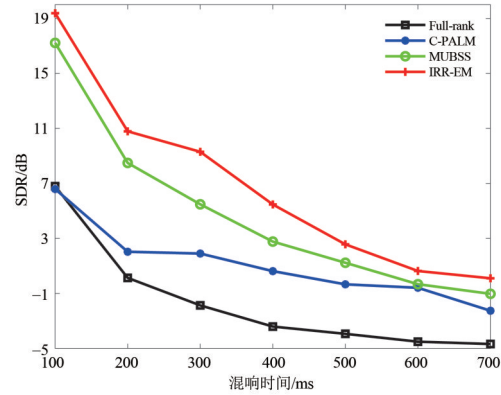
它们的值越大,说明盲源分离性能越好.

4.3 实验结果

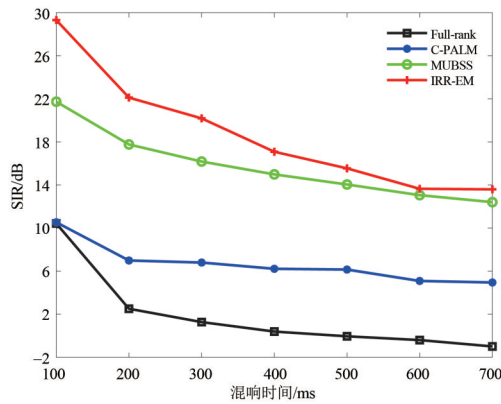
4.3.1 欠定情况下的盲源分离实验分析

实验 1 从 SiSEC'08 数据集中分别选取 3 组女性英文语音混叠信号 ($I=3, J=2$) 和 4 组女性英文语音混叠信号 ($I=4, J=2$), 源信号的分离性能分别如图 2、图 3 所示, 将 IRR-EM 算法与 Full-rank 算法^[13]、C-PALM 算法^[14] 和 MUBSS 算法^[30] 进行了对比, 实验结果表明, 随着混响时间的增加, SDR, SIR 和 SAR 值越来越小, 在高混响下 Full-rank 算法、C-PALM 算法和 MUBSS 算法随着混响时间增加逐渐失效, 而 IRR-EM 算法仍具有较好的分离效果, 说明 IRR-EM 算法可以有效地解决欠定情况下的盲源分离问题.

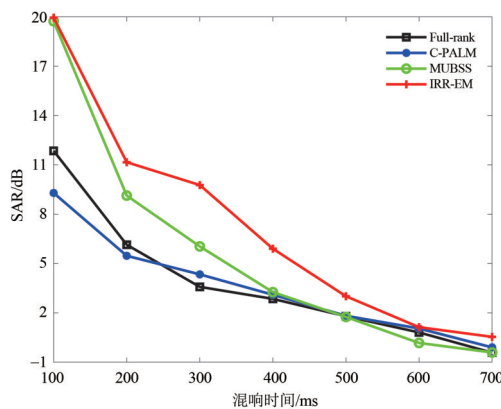
实验 2 考虑与实验 1 相同环境下的欠定中文语音



(a) SDR 随混响时间变化的关系



(b) SIR 随混响时间变化的关系

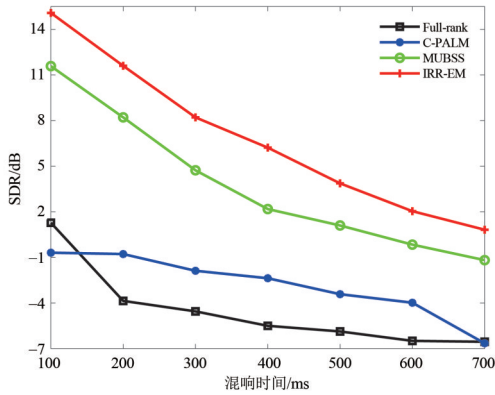


(c) SAR 随混响时间变化的关系

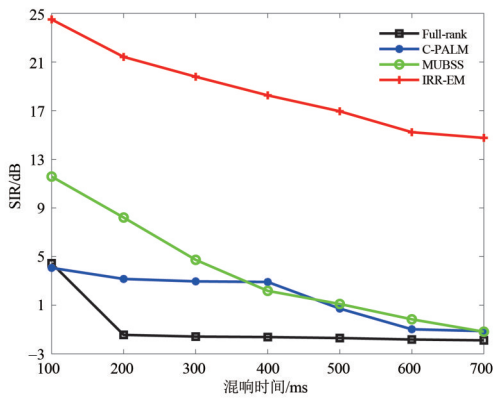
图 2 欠定混叠情况 3 组英文语音信号的盲源分离性能与混响时间的关系

混叠信号, 选取了 AISHELL 数据集中语音时长相同的中文语音信号, 即由两个麦克风接收 3 组和 4 组中文语音信号. 针对该欠定混叠情况的中文语音混合信号的

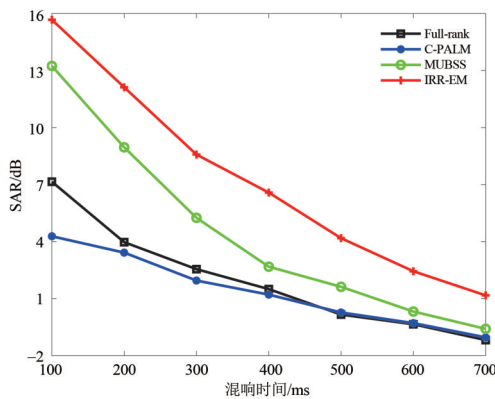
况下中文语音混合信号的盲源分离问题,不仅对低混响环境有效,还对高混响环境具有一定的自适应性.



(a) SDR 随混响时间变化的关系



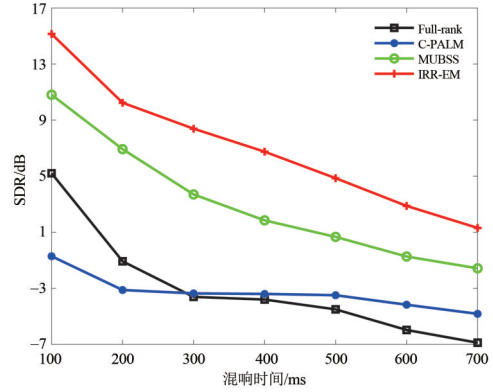
(b) SIR 随混响时间变化的关系



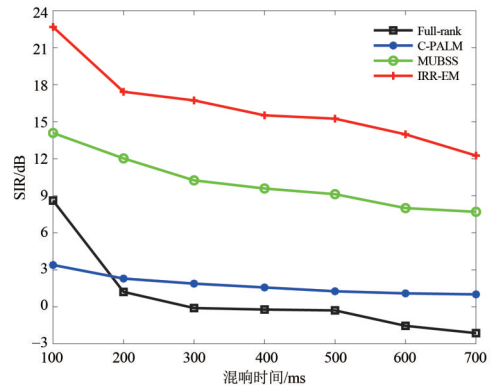
(c) SAR 随混响时间变化的关系

图 3 欠定混叠情况 4 组英文语音信号的盲源分离性能与混响时间的关系

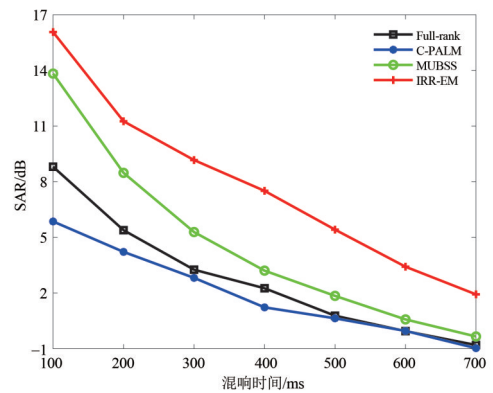
盲源分离,实验结果分别如图 4、图 5 所示,可见在高混响环境下,Full-rank 算法、C-PALM 算法和 MUBSS 算法的分离性能急剧下降,甚至失效.但 IRR-EM 算法的分离结果仍明显优于 Full-rank 算法、C-PALM 算法和 MUBSS 算法,说明 IRR-EM 算法可以有效地解决欠定情



(a) SDR 随混响时间变化的关系



(b) SIR 随混响时间变化的关系

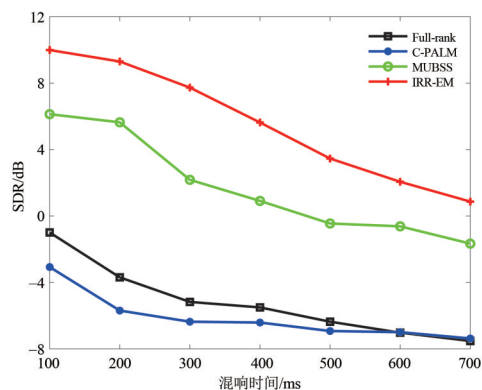


(c) SAR 随混响时间变化的关系

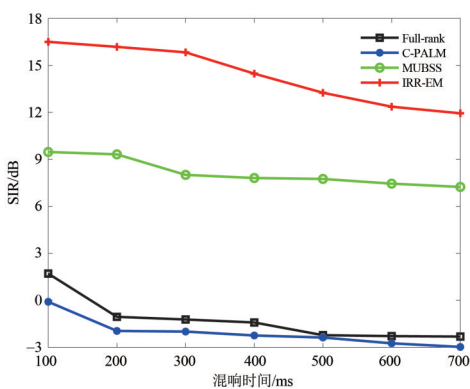
图 4 欠定混叠情况 3 组中文语音信号的盲源分离性能与混响时间的关系

4.3.2 鲁棒性分析

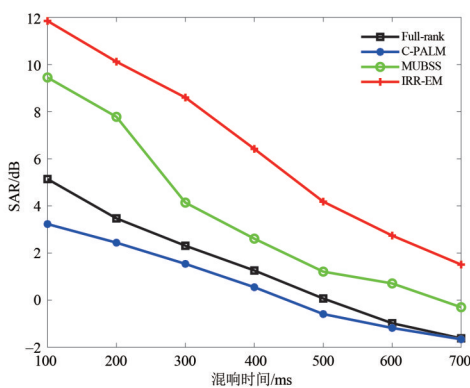
实验 3 为了验证 IRR-EM 算法对噪声的鲁棒性,分别选择实验 1 中的 3 组英文语音信号和实验 2 中的



(a) SDR 随混响时间变化的关系



(b) SIR 随混响时间变化的关系

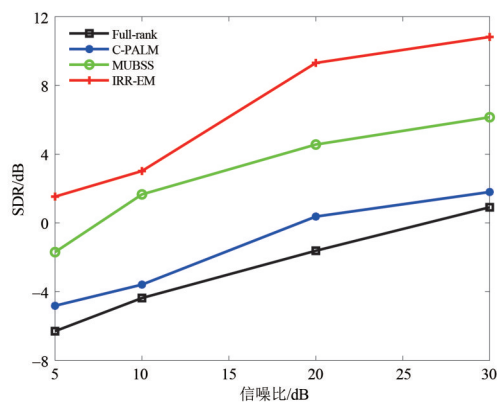


(c) SAR 随混响时间变化的关系

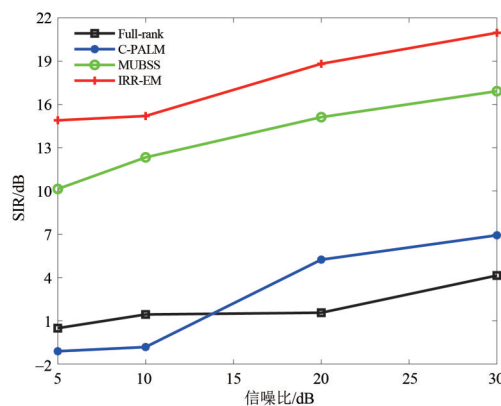
图5 欠定混叠情况4组中文语音信号的盲源分离性能与混响时间的关系

3组中文语音信号作为源信号,利用2个麦克风接收信号源,形成欠定的混合信号.考虑混响时间为250 ms,采用随机产生的高斯噪声进行实验,设置信噪比(Signal-Noise Ratio, SNR)从5 dB增加到30 dB.实验结果分别如图6、图7所示,可见随着SNR的增加,分

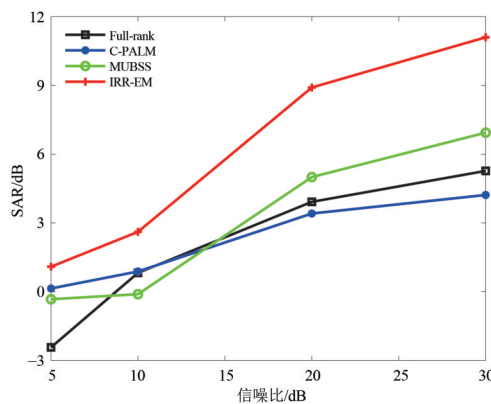
离性能指标 SDR, SIR 和 SAR 值逐渐上升并趋于稳定.在不同的信噪比下,相比 Full-rank 算法、C-PALM 算法和 MUBSS 算法,IRR-EM 算法都获得了更好的分离结果.这说明 IRR-EM 算法对噪声具有更好的鲁棒性.



(a) SDR 随信噪比变化的关系

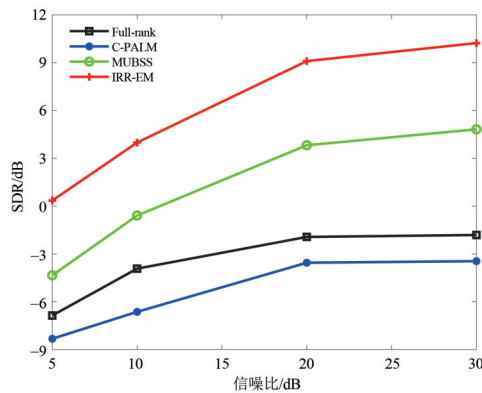


(b) SIR 随信噪比变化的关系

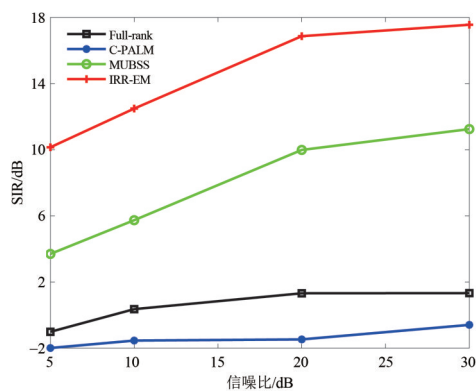


(c) SAR 随信噪比变化的关系

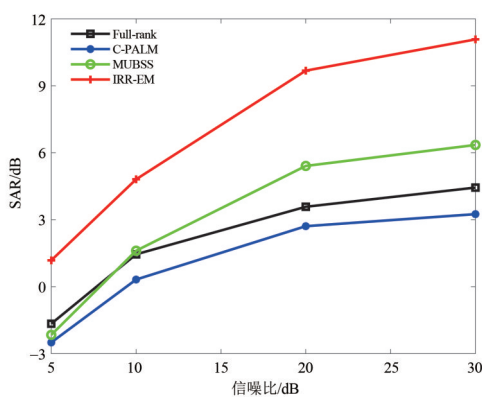
图6 含噪声的英文语音信号的盲源分离鲁棒性分析



(a) SDR随信噪比变化的关系



(b) SIR随信噪比变化的关系



(c) SAR随信噪比变化的关系

图7 含噪声的中文语音信号盲源分离鲁棒性分析

5 结论

针对多通道欠定卷积语音混合信号的盲源分离问题,本文设计了一种基于脉冲响应重塑技术的期望最大化盲源分离算法,通过对脉冲响应进行重塑,提高了接收信号的质量,进而设计新的期望最大化算法获得模型参数更新规则,结合脉冲响应重塑技术的期望最

大化重构源信号. 实验结果验证了 IRR-EM 算法对欠定情况下中英文语音混合信号的盲源分离的有效性,其不仅在低混响时间下分离性能优越,在高混响时间下仍保持较好的分离性能. 同时,相比流行的 Full-rank 算法、C-PALM 算法和 MUBSS 算法,本文设计的 IRR-EM 算法实现了更好的盲源分离结果,证实了 IRR-EM 算法的优越性,以及对噪声具有较好的鲁棒性.

值得注意的是,本文考虑的是多通道的盲源分离问题,要求声音源和麦克风的位置是不变的. 针对单通道的信号分离以及移动环境下的源信号分离问题,本文算法受到限制. 在未来科研工作中,可考虑利用深度学习探索单通道的盲源分离问题,以及移动源信号的盲源分离问题.

参考文献

- [1] BEE M A, MICHEYL C. The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?[J]. *Journal of Comparative Psychology*, 2008, 122(3): 235-251.
- [2] WOODS K J P, MCDERMOTT J H. Schema learning for the cocktail party problem[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(14): E3313-E3322.
- [3] JUTTEN C, HERAULT J. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture[J]. *Signal Processing*, 1991, 24(1): 1-10.
- [4] 皮磊, 朱磊, 郑翔, 等. 基于改进 Wave-U-Net 跳跃连接的盲源分离算法[J]. *信号处理*, 2022, 38(4): 835-843.
PI L, ZHU L, ZHENG X, et al. Blind source separation algorithm based on improved wave-U-net skip connection [J]. *Journal of Signal Processing*, 2022, 38(4): 835-843. (in Chinese)
- [5] XIE Y, XIE K, YANG Q Y, et al. Reverberant blind separation of heart and lung sounds using nonnegative matrix factorization and auxiliary function technique[J]. *Biomedical Signal Processing and Control*, 2021, 69: 102899.
- [6] 刘秋红, 许漫坤, 李天昀, 等. 基于互补对称滤波器的 APCMA 信号的盲分离算法[J]. *电子学报*, 2020, 48(12): 2394-2401.
LIU Q H, XU M K, LI T Y, et al. Blind separation of APCMA signal based on complementary symmetric filters[J]. *Acta Electronica Sinica*, 2020, 48(12): 2394-2401. (in Chinese)
- [7] 李帅, 刘宏清, 彭鹏, 等. 混响环境下基于卷积模型的欠定盲源分离[J]. *信号处理*, 2021, 37(4): 624-632.
LI S, LIU H Q, PENG P, et al. Underdetermined blind

- source separation based on convolution model in reverberant environment[J]. *Journal of Signal Processing*, 2021, 37(4): 624-632. (in Chinese)
- [8] SHI Y H, ZENG W M, WANG N Z, et al. A new constrained spatiotemporal ICA method based on multi-objective optimization for fMRI data analysis[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2018, 26(9): 1690-1699.
- [9] 田宝平, 应昊蓉, 杨文境, 等. 结合 ICA 和复数神经网络的双麦阵列盲源分离方法[J]. *信号处理*, 2021, 37(11): 2185-2192.
- TIAN B P, YING H R, YANG W J, et al. Blind source separation of binary array based on ICA and complex neural network[J]. *Journal of Signal Processing*, 2021, 37(11): 2185-2192. (in Chinese)
- [10] FENG F C, KOWALSKI M. Revisiting sparse ICA from a synthesis point of view: Blind Source Separation for over and underdetermined mixtures[J]. *Signal Processing*, 2018, 152: 165-177.
- [11] GEORGIEV P, THEIS F, CICHOCKI A. Sparse component analysis and blind source separation of underdetermined mixtures[J]. *IEEE Transactions on Neural Networks*, 2005, 16(4): 992-996.
- [12] XIE Y, XIE K, XIE S L. Underdetermined blind source separation of speech mixtures unifying dictionary learning and sparse representation[J]. *International Journal of Machine Learning and Cybernetics*, 2021, 12(12): 3573-3583.
- [13] DUONG N Q K, VINCENT E, GRIBONVAL R. Underdetermined reverberant audio source separation using a full-rank spatial covariance model[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010, 18(7): 1830-1840.
- [14] FENG F C, KOWALSKI M. Underdetermined reverberant blind source separation: Sparse approaches for multiplicative and convolutive narrowband approximation[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, 27(2): 442-456.
- [15] GUO J Y, LIU S, YU K, et al. An ultrahigh voltage shunt reactor acoustic signal separation method based on masking beamforming and underdetermined blind source separation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-8.
- [16] NION D, MOKIOS K N, SIDIROPOULOS N D, et al. Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010, 18(6): 1193-1207.
- [17] QIN G D, AMIN M G, ZHANG Y D. DOA estimation exploiting sparse array motions[J]. *IEEE Transactions on Signal Processing*, 2019, 67(11): 3013-3027.
- [18] SAWADA H, MUKAI R, ARAKI S, et al. A robust and precise method for solving the permutation problem of frequency-domain blind source separation[J]. *IEEE Transactions on Speech and Audio Processing*, 2004, 12(5): 530-538.
- [19] SU Q, ZHANG X W, SHA N, et al. Underdetermined blind direction-of-arrival estimation using a moving platform[J]. *IEEE Signal Processing Letters*, 2022, 29: 2532-2536.
- [20] 马宝泽, 张天骐, 安泽亮, 等. 基于张量分解的卷积盲源分离方法[J]. *通信学报*, 2021, 42(8): 52-60.
- MA B Z, ZHANG T Q, AN Z L, et al. Convolutive blind source separation method based on tensor decomposition [J]. *Journal on Communications*, 2021, 42(8): 52-60. (in Chinese)
- [21] XIE K, ZHOU G X, YANG J J, et al. Eliminating the permutation ambiguity of convolutive blind source separation by using coupled frequency bins[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(2): 589-599.
- [22] MA B Z, LI G J, YI C. Tensor-based underdetermined blind identification of instantaneous mixtures[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2023, 70(1): 346-350.
- [23] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 1999, 401(6755): 788-791.
- [24] 解元, 邹涛, 孙为军, 等. 面向卷积混叠环境下的盲源分离新方法[J]. *自动化学报*, 2023, 49(5): 1062-1072.
- XIE Y, ZOU T, SUN W J, et al. Novel blind source separation method for convolutive mixed environment[J]. *Acta Automatica Sinica*, 2023, 49(5): 1062-1072. (in Chinese)
- [25] WANG T H, YANG F R, YANG J. Convolutive transfer function-based multichannel nonnegative matrix factorization for overdetermined blind source separation[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2022, 30: 802-815.
- [26] 刘升东, 杨飞然, 杨军. 基于最小体积约束的频域卷积盲源

分离[J]. 信号处理, 2023, 39(5): 829-836.

LIU S D, YANG F R, YANG J. Frequency-domain convolutive blind source separation with minimum volume constraint[J]. Journal of Signal Processing, 2023, 39(5): 829-836. (in Chinese)

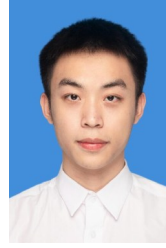
- [27] KOUNADES-BASTIAN D, GIRIN L, ALAMEDA-PINEDA X, et al. A variational EM algorithm for the separation of time-varying convolutive audio mixtures[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(8): 1408-1423.
- [28] AL-TMEME A, WOO W L, DLAY S S, et al. Underdetermined convolutive source separation using GEM-MU with variational approximated optimum model order NMF2D[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(1): 35-49.
- [29] XIE Y, XIE K, XIE S L. Underdetermined convolutive blind separation of sources integrating tensor factorization and expectation maximization[J]. Digital Signal Processing, 2019, 87: 145-154.
- [30] XIE K, JIANG K Y, YANG Q Y. Multi-channel underdetermined blind source separation for recorded audio mixture signals using an unmanned aerial vehicle[J]. IET Communications, 2021, 15(10): 1412-1422.
- [31] MERTINS A, MEI T M, KALLINGER M. Room impulse response shortening/reshaping with infinity and p -norm optimization[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(2): 249-259.
- [32] DAU T, KOLLMEIER B, KOHLRAUSCH A. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers[J]. The Journal of the Acoustical Society of America, 1997, 102(5): 2892-2905.
- [33] WINTER S, KELLERMANN W, SAWADA H, et al. MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and-norm minimization[J]. EURASIP Journal on Advances in Signal Processing, 2007: 024717.
- [34] VINCENT E, GRIBONVAL R, FEVOTTE C. Performance measurement in blind audio source separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1462-1469.

作者简介



解 元 男, 1989年出生于安徽利辛县. 现为广州大学机械与电气工程学院讲师、硕士生导师. 主要研究方向为盲信号分离、信号处理和机器学习等.

E-mail: xieyuan@gzhu.edu.cn



张 旭 男, 2000年出生于广东韶关市. 现为广州大学机械与电气工程学院硕士研究生. 主要研究方向为盲信号分离.

E-mail: sgzx8480@e.gzhu.cn



邹 涛(通讯作者) 男, 1975年出生于辽宁营口. 现为广州大学机械与电气工程学院教授、博士生导师. 主要研究方向为工业过程建模与仿真、模型预测控制、先进过程控制和实时优化技术研究与应用.

E-mail: tzou@gzhu.edu.cn



马 鸽 女, 1988年出生于山东济宁. 广州大学机械与电气工程学院副教授、硕士生导师. 主要研究方向为图像处理、机器视觉、控制理论与应用等.

E-mail: m_ge@gzhu.edu.cn



余锦视 男, 1990年出生于广东潮州市. 现为广州大学机械与电气工程学院博士后. 主要研究包括非负矩阵分解、张量分解、张量完成和机器学习等.

E-mail: jinshi.yu@gzhu.edu.cn



孙为军 男, 1975年出生于安徽马鞍山市. 现为广东工业大学自动化学院副教授、硕士生导师. 主要研究方向为模式识别、机器学习等.

E-mail: gdutswj@163.com