

面向高混响环境的欠定卷积盲源分离算法

解元¹, 邹涛¹, 孙为军², 谢胜利³

(1. 广州大学机械与电气工程学院, 广东 广州 510006; 2. 广东工业大学智能检测与制造物联教育部重点实验室, 广东 广州 510006;
3. 广东工业大学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: 为了解决高混响环境下欠定卷积混叠信号的分离问题, 提出一种新的欠定卷积盲源分离算法。针对高混响环境的影响, 设计全局脉冲响应网络削弱混响回声, 提高信号质量。基于全局脉冲响应网络建立新的时频域混叠信号数学模型, 采用全局脉冲响应矩阵缩短了传统脉冲响应的长度, 降低了高混响带来的模型变换近似误差。基于非负矩阵分解理论设计模型参数的实时更新学习规则, 将源信号分离问题转换为模型参数优化问题, 实现混叠信号的盲源分离。实验结果表明, 所提算法可以有效地实现中英文语音、音乐混叠信号的盲源分离, 与现有比较流行的盲源分离算法的对比较验证了所提算法的优越性。

关键词: 盲源分离; 欠定卷积混叠; 高混响环境; 全局脉冲响应网络; 非负矩阵分解

中图分类号: TN911.23

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023027

Algorithm of underdetermined convolutive blind source separation for high reverberation environment

XIE Yuan¹, ZOU Tao¹, SUN Weijun², XIE Shengli³

1. School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China
2. Key Laboratory of Intelligent Detection and the Internet of Things in Manufacturing, Ministry of Education, Guangdong University of Technology, Guangzhou 510006, China
3. Key Laboratory of Intelligent Information Processing and System Integration of IoT, Ministry of Education, Guangdong University of Technology, Guangzhou 510006, China

Abstract: To separate the underdetermined convolutive mixture signals in the high reverberation environment, a novel algorithm of underdetermined convolutive blind source separation was proposed. Aiming at the influence of high reverberation environment, a global impulse response network was designed to weaken reverberation echo, improving signal quality. A new mathematical model of time-frequency mixing signals was established based on the global impulse response network. The global impulse response matrix which shortened the length of the traditional impulse response, reduced the approximation error of model transformation caused by high reverberation. The real-time update learning rules of model parameters were designed based on the theory of nonnegative matrix factorization, and the source signal separation problem was converted into the model parameter optimization problem, achieving blind source separation of mixing signals. Experimental results show that the proposed algorithm can effectively realize the blind source separation of Chinese and English speech and music signals, and the comparison with existing popular algorithms verified the superiority of the proposed algorithm.

Keywords: blind source separation, underdetermined convolutive mixture, high reverberation environment, global impulse response network, nonnegative matrix factorization

收稿日期: 2022-10-10; 修回日期: 2022-12-26

通信作者: 邹涛, tzou@gzhu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62003095, No.52171331); 广东省重点领域研发计划基金资助项目 (No.2019B01054002)

Foundation Items: The National Natural Science Foundation of China (No.62003095, No.52171331), Research and Development Plan of Key Fields in Guangdong Province (No.2019B01054002)

0 引言

盲源分离, 又称盲信号分离, 是指仅根据接收的混叠信号(观测信号)分离或恢复未知源信号, 目的是求得源信号的最佳估计^[1-2]。其因强大的分离功能已在通信信号处理^[3]、生物医学信号处理(心肺音信号分离等)^[4]、图像处理^[5]以及语音信号处理^[6-7]等多领域得到了广泛的应用。在实际的信号接收过程中, 传感器的数目往往小于源信号的数目(即欠定混叠), 导致通道的盲辨识极具挑战性, 特别是在高混响复杂环境下, 可听回声对音质有着重要的影响, 导致接收信号具有复杂性, 给源信号的分离带来了巨大的挑战, 传统的盲源分离算法无法彻底解决该类问题。

对人类听觉系统特性的研究表明, 当回声低于由直接声音引起的掩蔽极限时, 将听不到回声, 这就是人类听觉系统的时间掩蔽效应^[8-9]。当回声刚好低于掩蔽极限时人类听觉系统是听不见的, 人类听觉系统对那些超过掩蔽极限的回声非常敏感。在低混响环境下, 可利用空间脉冲响应重塑技术完全消除可听回声, 而不影响声音的质量。然而, 在高混响环境下, 可听回声的存在是不可避免的, 对此, 一些改进的空间脉冲响应重塑技术被提出^[10-11]。其中, Jungmann 等^[10]结合声学多输入多输出信道串扰消除和空间脉冲响应重塑技术降低了混响的影响。Mertins 等^[11]提出的基于凸正则化技术实现了空间脉冲响应重塑和串扰消除。然而, 以上研究主要集中于空间脉冲响应重塑技术本身, 而基于该技术的盲源分离研究尚不多见。鉴于此, 本文在盲源分离算法中引入空间脉冲响应重塑技术, 通过改进该技术以及设计新的盲源分离算法, 提出面向高混响复杂环境的欠定卷积盲源分离算法。

目前, 比较流行的盲源分离算法主要利用短时傅里叶变化把时域混叠信号转换到频域中, 根据信号在频域上的统计特性^[12]、独立性^[13]、非负性^[14]、稀疏性^[15]等性质, 设计相应的时频域盲源分离算法。在模型变换过程中, 把时域上的卷积混叠变换成频域上的瞬时线性混叠, 减少了时域卷积计算带来的复杂性。在低混响混叠环境下, 该模型变换具有较低的近似误差, 由此衍生出一系列欠定卷积盲源分离算法^[16-19]。其中, 文献[16-17]基于联合矩阵块对角化方法, 实现了卷积混叠信号的盲源分离; 文献[18-19]基于矩阵状态协方差模型, 提出了迭代

期望最大化算法估计模型参数, 通过实时更新优化模型参数, 再用维纳滤波法分离源信号。然而, 这类算法收敛速度受限, 而且对通道阶数与算法初始化较敏感。

非负矩阵分解是一种机器学习算法框架, 将非负矩阵分解为 2 个低秩非负因子矩阵的乘积^[20], 在处理卷积混叠盲源分离问题中, 通过将源信号的功率谱密度矩阵分解为 2 个非负矩阵的乘积, 一系列基于非负矩阵分解源模型的盲源分离算法被提出^[21-23]。其中, Sawada 等^[21]基于非负矩阵分解的低秩源模型, 利用多元复高斯分布的统计模型定义多通道欧几里得距离和多通道 IS (Itakura-Saito) 散度, 为了最小化这种散度, 通过设计适当的辅助函数, 推导出乘法更新形式的优化算法; Sekiguchi 等^[22]假设源图像遵循无约束全秩空间协方差矩阵的多元复高斯分布, 将空间协方差矩阵限制为以频率方式联合对角化满秩矩阵, 从而实现了快速多通道非负矩阵分解; Wang 等^[23]通过最小体积先验分布来增强源模型的可识别性, 利用最小体积正则化多通道非负矩阵分解, 最大化分离源的后验分布, 保证了收敛的稳定性。

然而, 基于非负矩阵分解设计的优化算法对模型参数的初始化比较敏感, 限制了算法的自适应性。另外, 时域上的卷积混叠模型通过短时傅里叶变换变换到频域上的瞬时线性混叠模型是一种近似变换模型, 其成立的前提条件是短时傅里叶变换的窗长度远大于脉冲响应的长度, 而在高混响环境下该条件是很难得到满足的, 极易导致较大的模型近似误差, 所以在此变换模型下设计的算法往往不适用于高混响环境。对此, 文献[24-25]提出基于卷积传递函数的卷积窄带近似, 该卷积宽带近似模型能避免以上条件的限制, 更精准地近似时域上的卷积模型, 适用于高混响混叠情形, 但其是卷积计算, 容易带来更高的计算量。

鉴于目前研究现状, 面向高混响环境的欠定卷积盲源分离问题仍然存在以下难点。

- 1) 在高混响复杂环境下, 可听回声和混响对音质有着重要的影响, 导致接收信号具有复杂性。
- 2) 缺乏对高混响环境下欠定卷积混叠信号精确的数学建模, 导致模型近似误差增大。
- 3) 欠定卷积混叠盲源分离实质上是一个非线性问题, 其求解困难, 由于外界环境的复杂性, 导致传统盲源分离算法的性能受限。

针对以上问题, 本文从空间脉冲响应重塑的角度出发, 结合欠定卷积盲源分离研究思路, 提出一种面向高混响环境的欠定卷积盲源分离算法——全局脉冲响应欠定盲源分离 (GIR-UBSS, global impulse response underdetermined blind source separation)。本文创新点概括如下。

1) 设计了全局脉冲响应网络, 通过优化可调滤波器削弱可听回声的影响, 提高了信号的质量。

2) 构建了面向高混响复杂环境的时频域混叠信号数学模型, 降低了模型近似误差, 对高混响环境具有较好的自适应性。

3) 提出了一种 GIR-UBSS 算法, 设计了新模型下参数的实时更新规则, 实现了源信号的盲分离。理论分析与一系列仿真实验验证了 GIR-UBSS 算法的有效性与优越性。

1 问题描述

对高混响环境下记录的混叠信号进行如下数学建模。

$$x_j(t) = \sum_{i=1}^I \sum_{\tau=0}^{L-1} a_{ji}(t) s_i(t) + b(t) \quad (1)$$

其中, $x_j(t)$ 是第 j 个通道记录的混叠信号, 时间 t 是连续的, $i=1, 2, \dots, I$ 是源信号数目, $j=1, 2, \dots, J$ 是传感器数目, $a_{ji}(t)$ 是第 i 个源信号到第 j 个通道过程中产生的空间脉冲响应, $s_i(t)$ 是第 i 个源信号, τ 是时延, L 是脉冲响应的长度, $b(t)$ 是噪声。利用矩阵向量的形式, 式(1)可表示为

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) + \mathbf{b}(t) = \sum_{\tau=0}^{L-1} \mathbf{A}(\tau) \mathbf{s}(t - \tau) + \mathbf{b}(t) \quad (2)$$

其中, $*$ 是卷积符号, $\mathbf{x}(t) = [x_1(t), \dots, x_J(t)]^T$, $\mathbf{s}(t) = [s_1(t), \dots, s_I(t)]^T$, $\mathbf{b}(t) = [b_1(t), \dots, b_J(t)]^T$, $\mathbf{A}(\tau)$ 是混叠系统。本文考虑高混响环境下的欠定卷积混叠盲源分离问题, 即 $I > J$ 为欠定混叠, 且空间混响时间逐渐增大导致高混响。盲源分离的目的是仅根据接收的混叠信号 $\mathbf{x}(t)$ 分离源信号 $\mathbf{s}(t)$ 。

2 本文算法

2.1 全局脉冲响应网络的设计

在高混响环境下, 接收到的混叠信号常常伴随混响回声, 为了消除或削弱可听回声的影响, 本文设计一种全局脉冲响应网络, 如图 1 所示, 该网络设计思路来源于信道串扰消除和空间脉冲响应重塑技术^[10]。

考虑在接收器前安装 L 个扩音器, 其中, I 个源信号经过 L 个扩音器传递到 J 个接收器, h_{li} 是第 i 个源信号通过第 l 个扩音器产生的脉冲响应, 其长度为 L_h , A_{jl} 是第 l 个扩音器到达第 j 个接收器产生的空间脉冲响应, 长度为 L_a , 则从第 i 个源信号到达第 j 个接收器产生的全局脉冲响应可以表示为

$$\mathbf{g}_{ji} = \sum_{l=1}^L A_{jl} * h_{li} \quad (3)$$

其中, \mathbf{g}_{ji} 的长度为 $L_g = L_a + L_h - 1$ 。利用矩阵向量的形式表示为

$$\mathbf{g}_{ji} = \mathbf{A}_{jl} * \mathbf{h}_{li} \quad (4)$$

其中, $\mathbf{g}_{ji} = [g_{ji}(0), \dots, g_{ji}(L_g - 1)]^T$, \mathbf{A}_{jl} 是维数为 $L_g \times L_h$ 的卷积矩阵, $\mathbf{h}_{li} = [h_{li}(0), \dots, h_{li}(L_h - 1)]^T$ 。为了使全局脉冲响应 \mathbf{g}_{ji} 尽可能地达到期望的脉冲响应 \mathbf{d}_{ji} , 即

$$\mathbf{g}_i = \mathbf{A} \mathbf{h}_i \approx \mathbf{d}_i \quad (5)$$

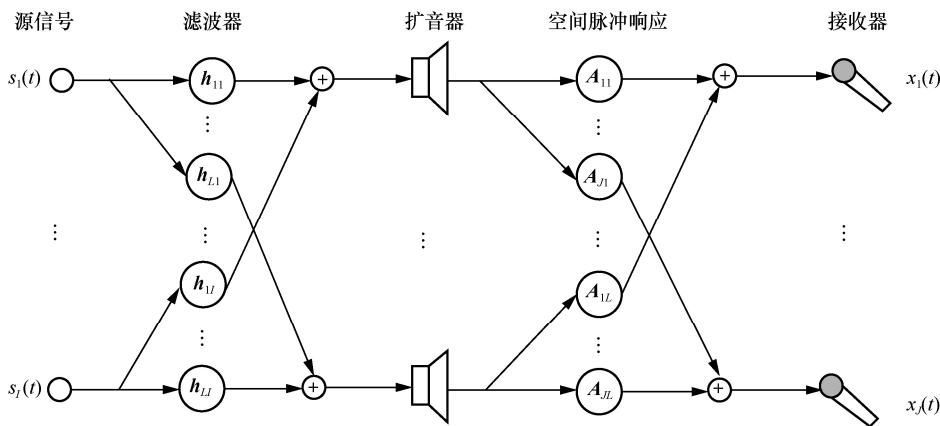


图 1 全局脉冲响应网络的设计流程

$$\text{其中, } \mathbf{g}_i = \begin{pmatrix} \mathbf{g}_{i1} \\ \vdots \\ \mathbf{g}_{ji} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1L} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{J1} & \cdots & \mathbf{A}_{JL} \end{pmatrix}, \mathbf{h}_i = \begin{pmatrix} \mathbf{h}_{i1} \\ \vdots \\ \mathbf{h}_{Li} \end{pmatrix},$$

$$\mathbf{d}_i = \begin{pmatrix} \mathbf{d}_{i1} \\ \vdots \\ \mathbf{d}_{ji} \end{pmatrix} \text{ 且 } \mathbf{d}_{ji} = \begin{pmatrix} \mathbf{d}_{ji}(0) \\ \vdots \\ \mathbf{d}_{ji}(L_g) \end{pmatrix}. \text{ 通过合理控制滤波器}$$

h_{ii} , 可以使全局脉冲响应 \mathbf{g}_{ji} 衰减到理想的结果, 从而消除可听回声, 达到对接收信号的清洗, 提高信号的清晰度和质量。

利用 2 个窗函数 \mathbf{W}_i^u 和 \mathbf{W}_i^d 得到不期望部分和期望部分, 分别为

$$\mathbf{g}_i^u = \mathbf{W}_i^u \mathbf{g}_i = \mathbf{W}_i^u \mathbf{A} \mathbf{h}_i \quad (6)$$

$$\mathbf{g}_i^d = \mathbf{W}_i^d \mathbf{g}_i = \mathbf{W}_i^d \mathbf{A} \mathbf{h}_i \quad (7)$$

定义

$$\mathbf{W}_i^u = \begin{bmatrix} \underbrace{0, 0, \dots, 0}_{N_1+N_2}, \underbrace{\mathbf{W}_0^T}_{N_3} \end{bmatrix}^T \quad (8)$$

$$\mathbf{W}_i^d = \begin{bmatrix} \underbrace{0, 0, \dots, 0}_{N_1}, \underbrace{1, 1, \dots, 1}_{N_2}, \underbrace{0, 0, \dots, 0}_{N_3} \end{bmatrix}^T \quad (9)$$

$$\mathbf{W}_0 = \frac{\beta - 1}{N_3 - 1} \mathbf{n} + 1 \quad (10)$$

其中, $N_1 = t_0 f_s$, $N_2 = \alpha f_s$, $N_3 = L_g - N_1 - N_2$, $n = 0, \dots, N_3 - 1$; α 和 β 是可调参数, 在设计滤波器过程中, 通过调节其值得到不同的窗函数, 实现不同的滤波目的。

最小化不期望部分同时最大化期望部分, 考虑优化问题

$$\min_{\mathbf{h}_i} \ln \left(\frac{\|\mathbf{g}_i^u\|_{p_u}}{\|\mathbf{g}_i^d\|_{p_d}} \right) = \ln \left(\frac{\|\mathbf{W}_i^u \mathbf{A} \mathbf{h}_i\|_{p_u}}{\|\mathbf{W}_i^d \mathbf{A} \mathbf{h}_i\|_{p_d}} \right) \quad (11)$$

其中,

$$\|\mathbf{g}_i^u\|_{p_u} = \left(\sum_{k=0}^{L_g-1} |\mathbf{g}_i^u(k)|^{p_u} \right)^{\frac{1}{p_u}} \quad (12)$$

$$\|\mathbf{g}_i^d\|_{p_d} = \left(\sum_{k=0}^{L_g-1} |\mathbf{g}_i^d(k)|^{p_d} \right)^{\frac{1}{p_d}} \quad (13)$$

其中, p_u 和 p_d 是可调的正整数, 通过调节其值设计不同范数的算法, 实现不同的脉冲响应重塑效果。

利用梯度下降法对式(11)求偏导, 得到

$$\begin{aligned} \nabla_{\mathbf{h}_i} \ln \left(\frac{\|\mathbf{g}_i^u\|_{p_u}}{\|\mathbf{g}_i^d\|_{p_d}} \right) &= \nabla_{\mathbf{h}_i} \left(\ln \|\mathbf{g}_i^u\|_{p_u} - \ln \|\mathbf{g}_i^d\|_{p_d} \right) = \\ &= \frac{1}{\|\mathbf{g}_i^u\|_{p_u}} \nabla_{\mathbf{h}_i} \|\mathbf{g}_i^u\|_{p_u} - \frac{1}{\|\mathbf{g}_i^d\|_{p_d}} \nabla_{\mathbf{h}_i} \|\mathbf{g}_i^d\|_{p_d} \end{aligned} \quad (14)$$

因此, 针对优化问题式(11)的学习规则为

$$\mathbf{h}_i^{\text{iter}+1} = \mathbf{h}_i^{\text{iter}} - \mu \left(\frac{1}{\|\mathbf{g}_i^u\|_{p_u}} \nabla_{\mathbf{h}_i} \|\mathbf{g}_i^u\|_{p_u} - \frac{1}{\|\mathbf{g}_i^d\|_{p_d}} \nabla_{\mathbf{h}_i} \|\mathbf{g}_i^d\|_{p_d} \right) \quad (15)$$

其中, iter 是迭代次数, μ 是学习率。

2.2 欠定卷积盲源分离算法的设计

2.2.1 模型变换

为解决模型式(2)下的盲源分离问题, 传统的方法利用短时傅里叶变换得到频域上近似的线性混叠模型, 表示为

$$\mathbf{x}_{fn} = \mathbf{A}_f \mathbf{s}_{fn} + \mathbf{b}_{fn} \quad (16)$$

其中, $f = 1, \dots, F$ 是频点指数, F 是短时傅里叶变换的窗长度, $n = 1, \dots, N$ 是时间窗指数,

$\mathbf{x}_{fn} = [x_{1,fn}, \dots, x_{J,fn}]^T$, $\mathbf{s}_{fn} = [s_{1,fn}, \dots, s_{I,fn}]^T$ 和

$\mathbf{b}_{fn} = [b_{1,fn}, \dots, b_{J,fn}]^T$ 分别是 $\mathbf{x}(t)$ 、 $\mathbf{s}(t)$ 和 $\mathbf{b}(t)$ 通过短时傅里叶变换得到的, 噪声 $\mathbf{b}(t)$ 是模拟现实生活中的真实噪声, $\mathbf{A}_f = [\mathbf{a}_{1f}, \dots, \mathbf{a}_{Jf}] \in \mathbb{C}^{J \times I}$ 是频域上的脉

冲响应混叠矩阵。然而, 这种变换成立的前提条件是短时傅里叶变换的窗长度 F 远大于脉冲响应的长度 L , 即 $F \gg L$ 。在高混响环境下, 随着混响时间 (RT, reverberation time) 的增加, 脉冲响应的长度 L 逐渐变大, 甚至超过窗长度 F , 导致线性混叠模型式(16)近似误差增大, 甚至无效。为了避免这种限制, 本文设计全局脉冲响应网络, 消除或削弱可听回声的影响, 缩短脉冲响应的长度, 降低高混响环境的影响, 建立如下近似模型。

$$\mathbf{x}_{fn} = \mathbf{G}_{fn} \mathbf{s}_{fn} + \mathbf{b}_{fn} \quad (17)$$

其中, $\mathbf{G}_{fn} = [\mathbf{g}_{1,fn}, \dots, \mathbf{g}_{J,fn}]$ 是全局脉冲响应混叠矩阵,

且 $p(\mathbf{g}_{:,fn} | \mathbf{g}_{:,fn-1}) = \mathbb{N}_c(\mathbf{g}_{:,fn}; \mathbf{g}_{:,fn-1}, \mathbf{R}_f^g)$,

$\mathbf{R}_f^g \in \mathbb{C}^{J \times J}$ 是协方差矩阵, $p(\mathbf{b}_{fn}) = \mathbb{N}_c(\mathbf{b}_{fn}; \mathbf{0}, n_f \mathbf{I}_J)$,

n_f 是衡量噪声的未知参数。该模型的优势是降低了高混响带来的模型变换近似误差。

2.2.2 非负矩阵分解模型

假设^[18]

$$\mathbf{s}_{fn} = \mathbf{M} \mathbf{c}_{fn} \quad (18)$$

其中, $s_{i,fn} = \sum_{k \in K_i} c_{k,fn}$, $\mathbf{M} \in \mathbb{N}^{I \times K}$, $M_{ik} = \begin{cases} 1, & k \in K_i \\ 0, & \text{其他} \end{cases}$,

其概率密度函数为

$$p(\mathbf{v}) = \mathbb{N}_c(\mathbf{c}_{fn}; \mathbf{0}; \text{diag}_K(u_{fn} \mathbf{v}_{kn})) \quad (19)$$

其中, $\mathbf{0}$ 是零向量, $\text{diag}(\cdot)$ 是对角化操作。因此, 源信号 \mathbf{s}_{fn} 的概率密度函数可写为

$$p(\mathbf{s}_{fn}) = \mathbb{N}_c\left(\mathbf{s}_{fn}; \mathbf{0}; \text{diag}_I\left(\sum_{k \in K_i} u_{fn} \mathbf{v}_{kn}\right)\right) \quad (20)$$

即

$$\mathbb{E}\left[|\mathbf{s}_i|^2\right] = \left\{ \mathbb{E}\left[|\mathbf{s}_{i,fn}|^2\right] \right\}_{fn} = \mathbf{U}_i \mathbf{V}_i \quad (21)$$

其中, $\mathbf{U}_i = \{u_{fk}\}_{f,k \in K_i}$ 和 $\mathbf{V}_i = \{v_{kn}\}_{k \in K_i, n}$ 均为非负矩阵。该非负矩阵分解模型通过提供谐波和非谐波模式模拟源信号, 其优势是避免源信号重构过程中出现的排序问题。

2.2.3 模型参数的更新规则

为了从观测数据集 $\{\mathbf{x}_{fn}\}_{f,n=1}^{F,N}$ 中估计变量 $\Phi = \{\mathbf{g}_{:,fn}, \mathbf{s}_{fn}, \mathbf{c}_{fn}\}_{f,n=1}^{F,N}$, 需要估计其对应的参数集 $\theta = \{\mathbf{R}_f^g, u_{fk}, v_{kn}, n_f\}_{f,n,k=1}^{F,N,K}$, 给定变量的后验分布^[18]

$$q(\Phi) = p(\Phi | \{\mathbf{x}_{fn}\}_{f,n=1}^{F,N}; \theta) \quad (22)$$

假设

$$q(\Phi) \approx \prod_{f=1}^F q(\mathbf{g}_{:,f1:N}) \prod_{f,n=1}^{F,N} q(\mathbf{s}_{fn}) \quad (23)$$

把式(18)代入式(23), 得到

$$q(\Phi) \approx \prod_{f=1}^F q(\mathbf{g}_{:,f1:N}) \prod_{f,n=1}^{F,N} q(\mathbf{M}\mathbf{c}_{fn}) \quad (24)$$

其子集 $\Phi_0 \subseteq \Phi$ 的最优后验分布可表示为

$$q(\Phi_0) \propto e^{\mathbb{E}_{q(\Phi_0)}[\ln p(\Phi_0 | \{\mathbf{x}_{fn}\}_{f,n=1}^{F,N}; \theta)]} \quad (25)$$

因此, 根据式(25), $\mathbf{g}_{:,fn}$ 的后验分布可写为

$$q(\mathbf{g}_{:,f1:N}) \propto p(\mathbf{g}_{:,f1:N}) \prod_{n=1}^N e^{\mathbb{E}_{q(\mathbf{s}_{fn})}[\ln p(\mathbf{x}_{fn} | \mathbf{g}_{:,fn}; \mathbf{s}_{fn})]} \quad (26)$$

即

$$q(\mathbf{v}) \propto p(\mathbf{g}_{:,f1:N}) \prod_{n=1}^N e^{\mathbb{E}_{q(\mathbf{M}\mathbf{c}_{fn})}[\ln p(\mathbf{x}_{fn} | \mathbf{g}_{:,fn}; \mathbf{M}\mathbf{c}_{fn})]} \quad (27)$$

把式(26)中的指数项简化为复高斯分布, 即

$$\mathbb{E}_{q(\mathbf{s}_{fn})}[\ln p(\mathbf{x}_{fn} | \mathbf{g}_{:,fn}; \mathbf{s}_{fn})] \propto \mathbb{N}_c(\boldsymbol{\mu}_{fn}^{sg}; \mathbf{g}_{:,fn}; \mathbf{R}_{fn}^{sg}) \quad (28)$$

其中, $\boldsymbol{\mu}_{fn}^{sg} \in \mathbb{C}^{JI}$ 和 $\mathbf{R}_{fn}^{sg} \in \mathbb{C}^{JI \times JI}$ 分别定义为

$$\boldsymbol{\mu}_{fn}^{sg} = \text{vec}(\mathbf{x}_{fn} \hat{\mathbf{s}}_{fn}^H (\mathbf{Q}_{fn}^{ns})^{-1}) \quad (29)$$

$$\mathbf{R}_{fn}^{sg} = (\mathbf{Q}_{fn}^{ns})^{-1} \otimes (n_f \mathbf{I}_J) \quad (30)$$

其中, vec 表示向量化操作, \otimes 表示 Kronecker 矩阵乘积, $\hat{\mathbf{s}}_{fn}^H \in \mathbb{C}^I$ 和 $\mathbf{Q}_{fn}^{ns} \in \mathbb{C}^{I \times I}$ 分别表示为

$$\hat{\mathbf{s}}_{fn}^H = \mathbf{M} \mathbb{E}_{q(\mathbf{c}_{fn})}[\mathbf{c}_{fn}] = \mathbf{M} \boldsymbol{\mu}_{fn}^{nc} \quad (31)$$

$$\mathbf{Q}_{fn}^{ns} = \mathbf{M} \mathbb{E}_{q(\mathbf{c}_{fn})}[\mathbf{c}_{fn} \mathbf{c}_{fn}^H] \mathbf{M}^T = \mathbf{M} \mathbf{R}_{fn}^{nc} \mathbf{M}^T + \hat{\mathbf{s}}_{fn} \hat{\mathbf{s}}_{fn}^H \quad (32)$$

其中, $\boldsymbol{\mu}_{fn}^{nc} \in \mathbb{C}^K$, $\mathbf{R}_{fn}^{nc} \in \mathbb{C}^{K \times K}$ 。

把式(28)代入式(26)中, 得到

$$q(\mathbf{g}_{:,f1:N}) \propto p(\mathbf{g}_{:,f1:N}) \prod_{n=1}^N \mathbb{N}_c(\boldsymbol{\mu}_{fn}^{sg}; \mathbf{g}_{:,fn}; \mathbf{R}_{fn}^{sg}) \quad (33)$$

为了更好地分析, 通过式(34)和式(35)表示边际分布和成对联合后验分布, 即

$$q(\mathbf{g}_{:,fn}) = \mathbb{N}_c(\mathbf{g}_{:,fn}; \boldsymbol{\mu}_{fn}^{ng}; \mathbf{R}_{fn}^{ng}) \quad (34)$$

$$q(\mathbf{g}_{:,fn}; \mathbf{g}_{:,fn-1}) = \mathbb{N}_c\left(\left[\begin{matrix} \mathbf{g}_{:,fn}^T \\ \mathbf{g}_{:,fn-1}^T \end{matrix}\right]^T; \boldsymbol{\mu}_{fn}^{eg}; \mathbf{R}_{fn}^{eg}\right) \quad (35)$$

其中, $\boldsymbol{\mu}_{fn}^{ng} \in \mathbb{C}^{JI}$ 和 $\mathbf{R}_{fn}^{ng} \in \mathbb{C}^{JI \times JI}$ 是边际分布的均值和协方差, $\boldsymbol{\mu}_{fn}^{eg} \in \mathbb{C}^{2JI}$ 和 $\mathbf{R}_{fn}^{eg} \in \mathbb{C}^{2JI \times 2JI}$ 是成对联合后验分布的均值和协方差。同理, 根据式(25), \mathbf{c}_{fn} 的后验分布可写为

$$q(\mathbf{c}_{fn}) \propto \sum_{k=1}^K p(c_{k,fn}) e^{\mathbb{E}_{q(\mathbf{g}_{:,fn})}[\ln p(\mathbf{x}_{fn} | \mathbf{g}_{:,fn}; \mathbf{c}_{fn})]} \quad (36)$$

用一个多元复高斯函数表示为

$$q(\mathbf{c}_{fn}) = \mathbb{N}_c(\mathbf{c}_{fn}; \boldsymbol{\mu}_{fn}^{nc}; \mathbf{R}_{fn}^{nc}) \quad (37)$$

其中, 后验均值 $\boldsymbol{\mu}_{fn}^{nc} \in \mathbb{C}^K$ 和协方差 $\mathbf{R}_{fn}^{nc} \in \mathbb{C}^{K \times K}$ 分别表示为

$$\boldsymbol{\mu}_{fn}^{nc} = \mathbf{R}_{fn}^{nc} \mathbf{M}^T (\boldsymbol{\mu}_{fn}^{ng})^H \frac{\mathbf{x}_{fn}}{n_f} \quad (38)$$

$$\mathbf{R}_{fn}^{nc} = \left[\text{diag}_K(u_{fk} v_{kn})^{-1} + \frac{1}{n_f} \mathbf{M}^T \mathbf{Q}_{fn}^T \mathbf{M} \right]^{-1} \quad (39)$$

其中, $\mathbf{Q}_{fn} \in \mathbb{C}^{I \times I}$, 其第 ir 个输入为 $[\mathbf{Q}_{fn}]_{ir} = \text{tr}\{\mathbf{Q}_{ir,fn}^{ng}\}$, 且 $\mathbf{Q}_{ir,fn}^{ng}$ 是滤波器的后二阶矩 \mathbf{Q}_{fn}^{ng} 的第 ir 个 $J \times J$ 方块。 \mathbf{Q}_{fn}^{ng} 可表示为

$$\mathbf{Q}_{fn}^{ng} = \mathbb{E}_{q(\mathbf{g}_{:,fn})}[\mathbf{g}_{:,fn} \mathbf{g}_{:,fn}^T] = \mathbf{R}_{fn}^{ng} + \boldsymbol{\mu}_{fn}^{ng} (\boldsymbol{\mu}_{fn}^{ng})^H \quad (40)$$

因此, 通过上述计算 \mathbf{g} 和 \mathbf{c} 的后验统计量, 可最大化对数似然, 表示为

$$\begin{aligned} \Gamma(\theta) = & \mathbb{E}_{q(\theta)} \ln p(\Phi, \{\mathbf{x}_{fn}\}_{f,n=1}^{F,N}; \theta) = \\ & \sum_{f,n=1}^{F,N} \mathbb{E}_{q(\mathbf{g}_{:,fn})q(s_{fn})} \left[\ln \mathbb{N}_c(\mathbf{x}_{fn}; \mathbf{g}_{fn} \mathbf{s}_{fn}, n_f \mathbf{I}_J) \right] + \sum_{f,n=1}^{F,N} \mathbb{E}_{q(c_{fn})} \left[\ln \mathbb{N}_c(c_{fn}; \mathbf{0}, \text{diag}_K(u_{fk} v_{kn})) \right] + \\ & \sum_f \left(\sum_{n=1}^{N-1} \mathbb{E}_{q(\mathbf{g}_{:,fn+1})} \left[\ln \mathbb{N}_c(\mathbf{g}_{:,fn+1}; \mathbf{g}_{:,fn}, \mathbf{R}_f^g) \right] + \mathbb{E}_{q(\mathbf{g}_{:,f1})} \left[\ln \mathbb{N}_c(\mathbf{g}_{:,f1}; \boldsymbol{\mu}_f^g, \mathbf{R}_f^g) \right] \right) \end{aligned} \quad (41)$$

通过优化式(41), 获得模型参数集 $\theta = \{\mathbf{R}_f^g, u_{fk}, v_{kn}, n_f\}_{f,n,k=1}^{F,N,K}$ 中的每个参数更新规则, 即

$$\mathbf{R}_f^g = \frac{1}{N} \left(\mathbf{Q}_{11,f}^{\xi g} - \mathbf{Q}_{12,f}^{\xi g} - \mathbf{Q}_{21,f}^{\xi g} + \mathbf{Q}_{22,f}^{\xi g} + \mathbf{R}_{f1}^{\eta g} \right) \quad (42)$$

$$u_{fk} = \frac{1}{N} \sum_{n=1}^N \frac{\mathbf{Q}_{kk,fn}^{\eta c}}{v_{kn}} \quad (43)$$

$$v_{kn} = \frac{1}{F} \sum_{f=1}^F \frac{\mathbf{Q}_{kk,fn}^{\eta c}}{u_{kn}} \quad (44)$$

$$n_f = \frac{1}{NJ} \sum_{n=1}^N \left(\mathbf{x}_{fn}^H \mathbf{x}_{fn} + \text{tr}(\mathbf{Q}_{fn}^T \mathbf{Q}_{fn}^{\eta s}) - 2 \text{Re} \left\{ \mathbf{x}_{fn}^H \boldsymbol{\mu}_{fn}^{\eta g} \hat{\mathbf{s}}_{fn} \right\} \right) \quad (45)$$

其中, $\mathbf{Q}_{fn}^{\xi g} = \sum_{n=1}^{N-1} \left(\mathbf{R}_{fn}^{\xi g} + \boldsymbol{\mu}_{fn}^{\xi g} (\boldsymbol{\mu}_{fn}^{\xi g})^H \right)$, $\mathbf{Q}_{fn}^{\eta c} = \mathbf{R}_{fn}^{\eta c} + \boldsymbol{\mu}_{fn}^{\eta g} (\boldsymbol{\mu}_{fn}^{\eta g})^H$, $\text{tr}(\cdot)$ 是求矩阵的迹, $\text{Re}(\cdot)$ 是求实部操作。通过以上更新规则对模型参数进行实时更新, 获得最优的模型参数。

2.2.4 频域源信号的估计

通过实时更新模型参数, 得到频域上的源信号表达式为

$$\hat{\mathbf{s}}_{fn} = \left[\text{diag}_I \left(\frac{1}{\sum_{k \in K_f} u_{fk} v_{kn}} \right) + \frac{\mathbf{Q}_{fn}}{n_f} \right]^{-1} (\boldsymbol{\mu}_{fn}^{\eta g})^H \frac{\mathbf{x}_{fn}}{n_f} \quad (46)$$

综上, 通过式(46)获得频域上的源信号, 再利用短时傅里叶逆变换得到时域上的源信号, 实现盲源分离。

3 实验

3.1 实验参数设置与评价准则

为了模拟真实环境, 利用国际上公用的模拟环境方法^[26], 创建一个有限脉冲响应房间, 该房间的维数是 $5 \text{ m} \times 3 \text{ m} \times 2.5 \text{ m}$, 固定 2 个传感器, 其坐标分别为 $[3 \ 1 \ 1.6]$ 和 $[3 \ 1.05 \ 1.6]$, 把源信号放置在 3 个位置 $[2 \ 0.5 \ 1.6]$ 、 $[2 \ 1 \ 1.6]$ 和 $[2 \ 1.5 \ 1.6]$, 即 3 个源信号两通道的欠定混叠 ($I=2, J=3$)。RT 设置为

100~900 ms, 值越大说明混响程度越强, 通过此模拟环境产生有混响的欠定混叠信号。在参数设置方面, 全局脉冲响应网络中 $L_a = L_h = f_s \text{ RT}$ (f_s 为信号的采样频率), $\alpha = 0.05$, $\beta = 2.0$, $p_u = 10$, $p_d = 20$, $\text{iter} = 1000$, $\mu = 10^{-6}$ 。在非负矩阵分解源模型中, 设置 $K_i = 20$; 在参数初始化方面, $\mathbf{R}_f^g = \mathbf{I}$, u_{fk} 和 v_{kn} 利用 KL (Kullback-Leibler) 散度

的非负矩阵分解获得, 且 $n_f = \frac{0.01 \sum_{fn} |x_{i,fn}|}{JN}$ 。

为了评价 GIR-UBSS 算法的有效性, 利用国际公认的评价准则: 信号失真比 (SDR, source-to-distortion ratio)、信号干扰比 (SIR, source-to-interference ratio)、信号伪像比 (SAR, source-to-artifacts ratio)^[27], 其值越大, 盲源分离性能就越好。因此, 利用评价准则 SDR、SIR、SAR 值的大小衡量盲源分离的好坏。

为了评价 GIR-UBSS 算法的优越性, 对比目前国际上比较流行的几种盲源分离算法: 变分期望最大化 (VEM, variational expectation-maximization) 算法^[19]、卷积近端交替线性化最小化 (C-PALM, convolutive proximal alternating linearized minimization) 算法^[24]、带正则化的窄带优化 (N-Regu, narrowband optimization with regularization) 算法^[24]、全秩空间协方差模型 (FullRankSCM, full-rank spatial covariance model) 算法^[18]。其中, VEM 算法先利用卡尔曼平滑器估计混叠矩阵和源信号参数, 再用维纳滤波法分离源信号。C-PALM 算法是一种卷积近似交替线性化极小化方法, 通过利用卷积窄带近似获得更好的模型近似, 减少了卷积核的长度, 避免了短时傅里叶变换窗函数长度的限制。N-Regu 算法是一种经典的带正则化的窄带优化方法, 利用了传统的线性窄带近似及 1-范数正则化。FullRankSCM 算法利用满秩空间协方差模型, 是 EM 算法中比较成熟的方法之一。以上对比算法是解决欠定卷积混叠盲源分离问题中比较流行的算法, 通过与这些流行的盲源分离

算法进行对比,可以很好地说明 GIR-UBSS 算法的优越性。同时,所有对比算法中的参数及模型采用的是与本文实验相同的设置,这样对比更有说服力。

3.2 全局脉冲网络去混响效果分析

为了定量描述全局脉冲网络实现的去混响效果,将可感知混响量化(nPRQ, perceivable reverberation quantization)度量作为评价准则^[11],当脉冲响应被完全重塑或没有时间系数超过时间掩蔽极限时, nPRQ=0, 说明混响被完全消除; 否则, nPRQ 越大, 脉冲响应被重塑的效果越差。为了验证全局脉冲网络对不同混响的影响, 设定混响时间 RT 为 100~900 ms, 并与原始脉冲网络进行对比, 实验结果如图 2 所示。相比于原始脉冲网络, 在低混响下, 全局脉冲网络可以完全消除混响, 使 nPRQ 趋于 0; 在高混响下, 全局脉冲网络可以削弱混响的影响, 提高信号的质量。

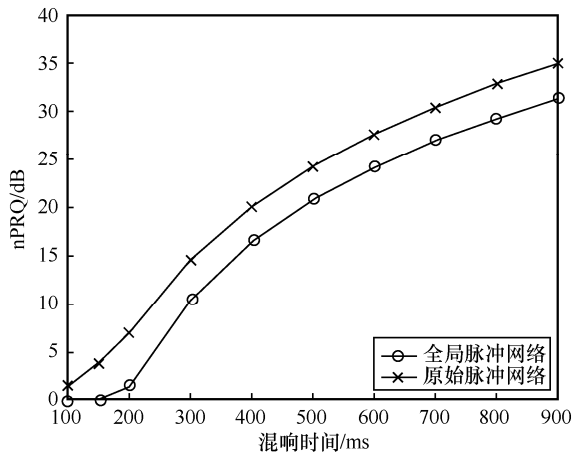


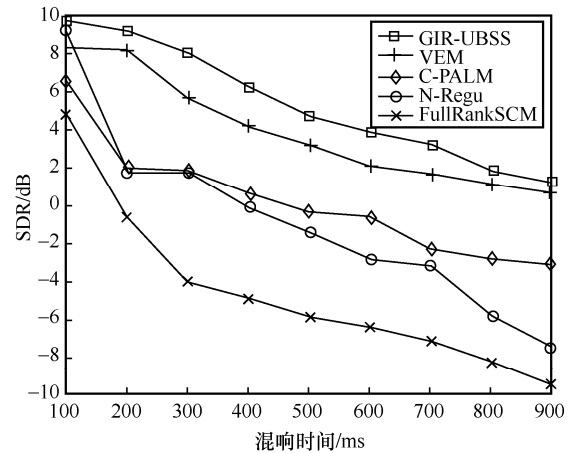
图 2 全局脉冲网络去混响效果

3.3 仿真实验 1: 英文语音信号欠定卷积盲源分离

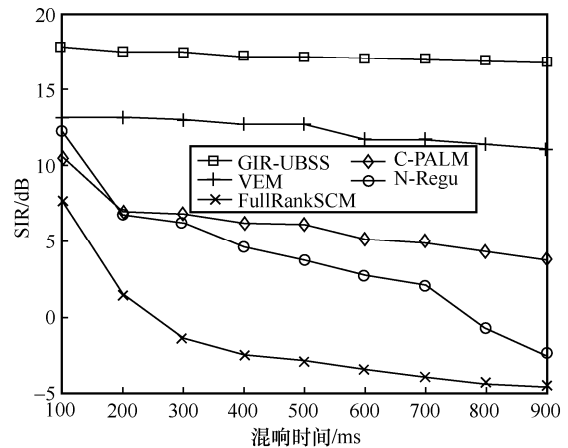
为了验证 GIR-UBSS 算法的有效性和优越性, 首选 3 组英文语音信号, 如表 1 所示。数据集来源于国际上公开的信号分离评价实验数据。本文创建一个两通道三语音源的欠定卷积混叠, 利用 GIR-UBSS 算法进行分离, 同时与几种比较流行的盲源分离算法进行对比, 实验结果如图 3 所示。从图 3 可知, 随着 RT 的增加, 算法的分离性能下降, 这是由于 RT 的增加带来了混响的复杂性, 导致分离越来越困难, 其中, SDR 和 SAR 的数值下降比较明显, 但是 SIR 的值比较稳定, 这是由于本文所设计的全局脉冲响应网络可以很好地减弱 RT 的影响, 从而得到稳定的 SIR。

表 1 仿真实验 1: 3 组英文语音

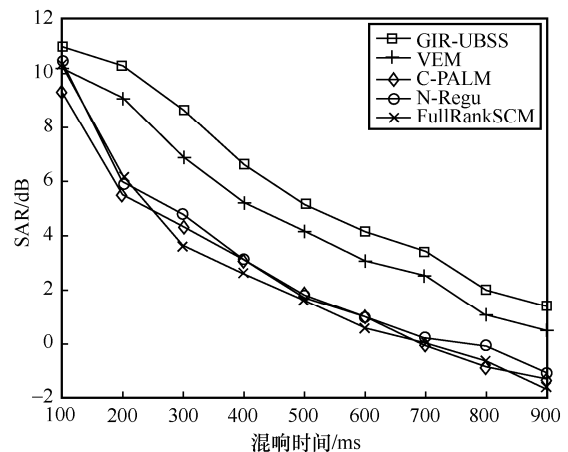
数据集	源信号	频率/kHz	时长/s
英文语音 1	dev1_female3_src_1	16	10
英文语音 2	dev1_female3_src_2	16	10
英文语音 3	dev1_female3_src_3	16	10



(a) SDR性能对比



(b) SIR性能对比



(c) SAR性能对比

图 3 英文语音信号欠定卷积盲源分离性能对比

在高混响环境下，对比算法分离性能下降严重，甚至失效，而 GIR-UBSS 算法依然可获得较好的分离结果。与对比算法中最好的分离结果相比，GIR-UBSS 算法得到的 SDR、SIR 和 SAR 值分别提高了约 1 dB、5 dB 和 1 dB，该实验验证了 GIR-UBSS 算法在高混响环境下分离欠定卷积语音混叠信号的有效性和优越性。

另外，为了从视觉的角度解析英文语音混叠信号分离情况，可视化混响时间为 300 ms 下的原始英文语音混叠信号波形以及分离后的英文语音信号波形，如图 4 所示。

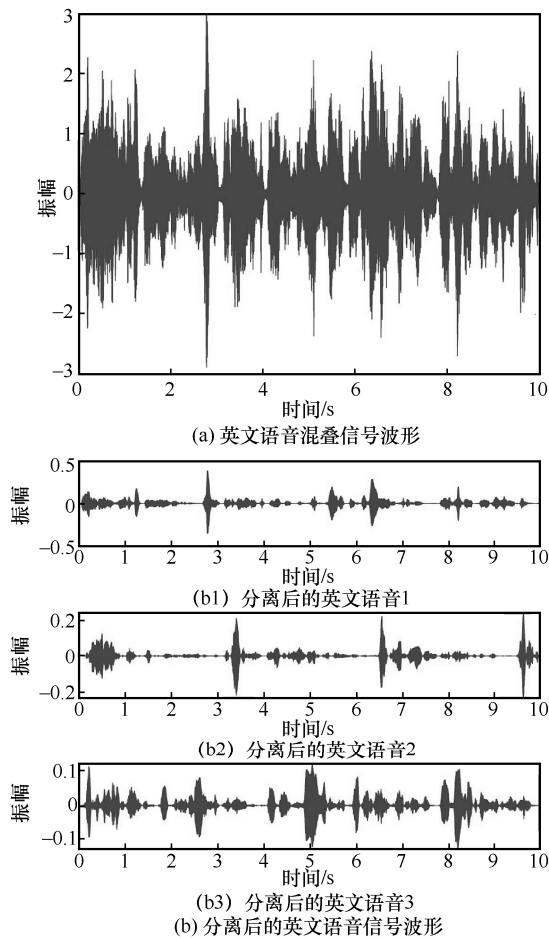


图 4 英文语音混叠信号波形以及分离后的英文语音信号波形

3.4 仿真实验 2：音乐信号欠定卷积盲源分离

为了验证 GIR-UBSS 算法对音乐混叠信号的有效性和优越性，测试 3 组音乐欠定卷积混叠信号，选取的源信号如表 2 所示，该音乐信号包括吉他声、人声和鼓声，全时长为 25 s，采样频率为 44.1 kHz，为了减少分离所需的时间，本文实验将信号截断为 10 s，下采样至 16 kHz，保持与实验 1 的一致性，盲源分离结果如图 5 所示。

表 2 仿真实验 2：3 组音乐信号

数据集	源信号	频率/kHz	时长/s
音乐信号 1	dev2_another_dreamer-the_ones_we_love_snip_69_94_guitar	44.1	25
音乐信号 2	dev2_another_dreamer-the_ones_we_love_snip_69_94_vocals	44.1	25
音乐信号 3	dev2_another_dreamer-the_ones_we_love_snip_69_94_drums	44.1	25

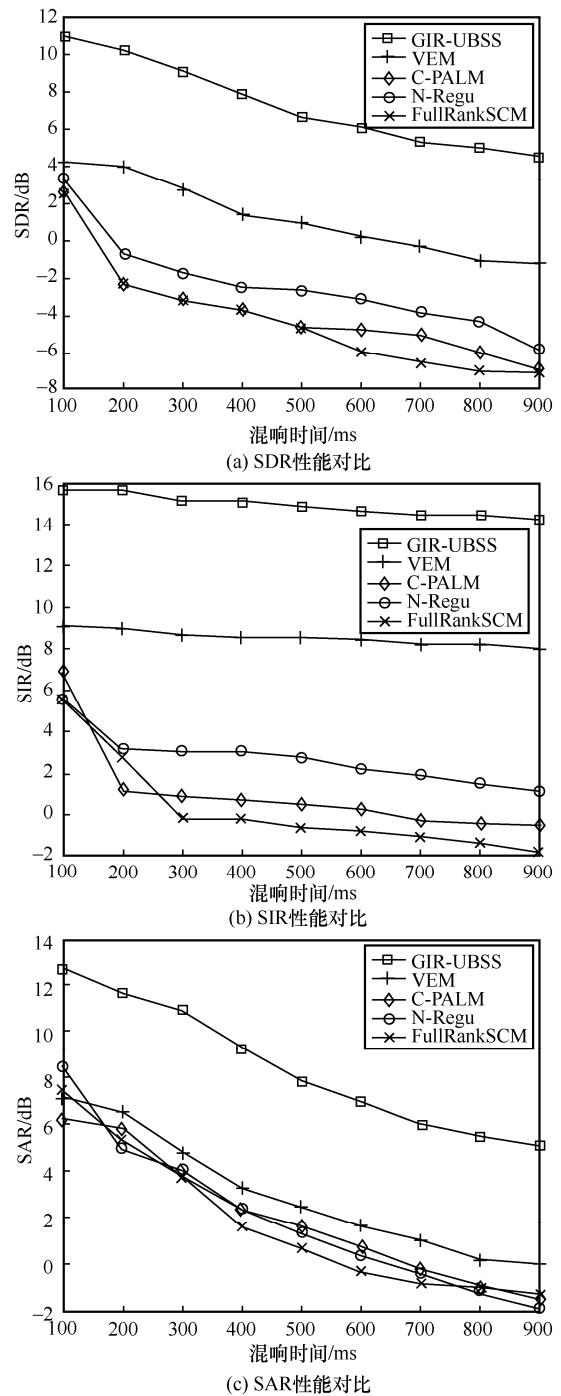


图 5 音乐信号欠定卷积盲源分离性能对比

从图 5 可知，随着 RT 的增大，算法的分离性能与预期基本一致。与对比算法中最好的分离结果

相比, GIR-UBSS 算法得到的 SDR、SIR、SAR 值分别提高了约 6 dB、7 dB 和 6 dB。该实验验证了 GIR-UBSS 算法在高混响环境下分离欠定卷积音乐混叠信号的有效性和优越性。

与仿真实验 1 的结果对比可发现, GIR-UBSS 算法在分离音乐混叠信号表现出更好的优越性。这是由于本文利用了非负矩阵分解源模型, 相比于语音源, 音乐源可以被较小数目的源成分表示, 更适用于非负矩阵分解模型, 从而获得更好的分离结果。

为了从视觉的角度解析音乐混叠信号分离情况, 可视化 300 ms 下的原始音乐混叠信号波形以及分离后的音乐信号波形, 如图 6 所示。

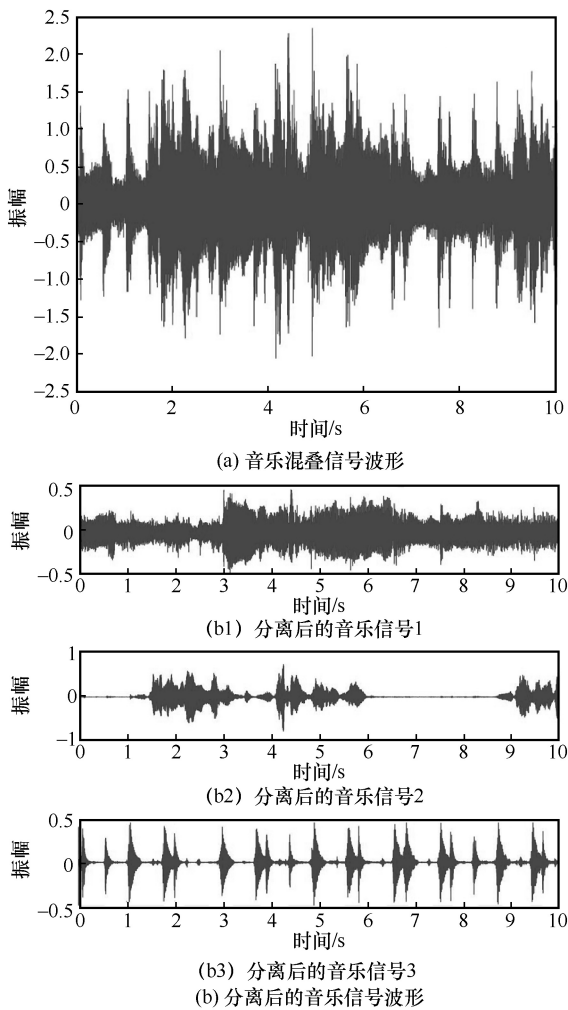


图 6 音乐混叠信号波形以及分离后的音乐信号波形

3.5 仿真实验 3: 中文语音信号欠定卷积盲源分离

为了验证 GIR-UBSS 算法对中文语音混叠信号的有效性和优越性, 测试 3 组中文语音混叠信号, 数据集来自国内公共中文语音数据集。选取 3 组语音信号, 如表 3 所示, 盲源分离结果如图 7 所示。与对比

算法中最好的分离结果相比, GIR-UBSS 算法得到的 SDR、SIR、SAR 值分别提高了约 1 dB、5 dB 和 1 dB。验证了 GIR-UBSS 算法对分离中文语音欠定卷积混叠信号仍然具有较好的有效性和优越性。

表 3 仿真实验 3: 3 组中文语音

数据集	源信号	频率/kHz	时长/s
中文语音 1	IC0936W0131	16	5
中文语音 2	IC0936W0134	16	5
中文语音 3	IC0936W0137	16	5

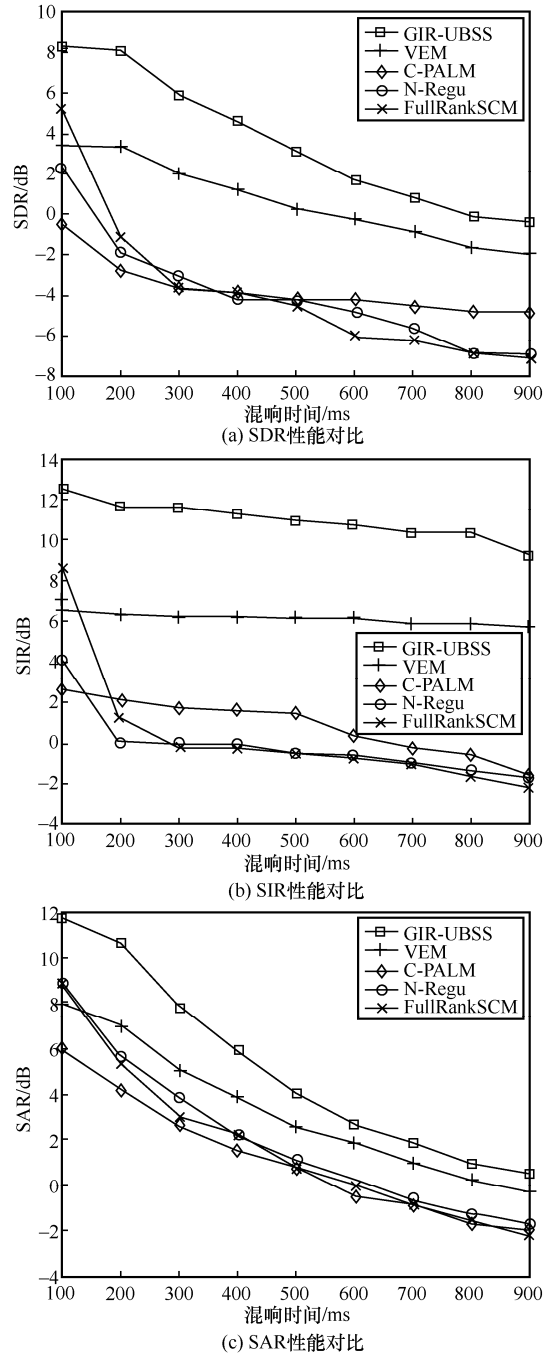


图 7 中文语音信号欠定卷积盲源分离性能对比

另外，为了从视觉的角度解析中文语音混叠信号分离情况，可视化混响时间为 300 ms 下的原始中文语音混叠信号波形以及分离后的中文语音信号波形如图 8 所示。

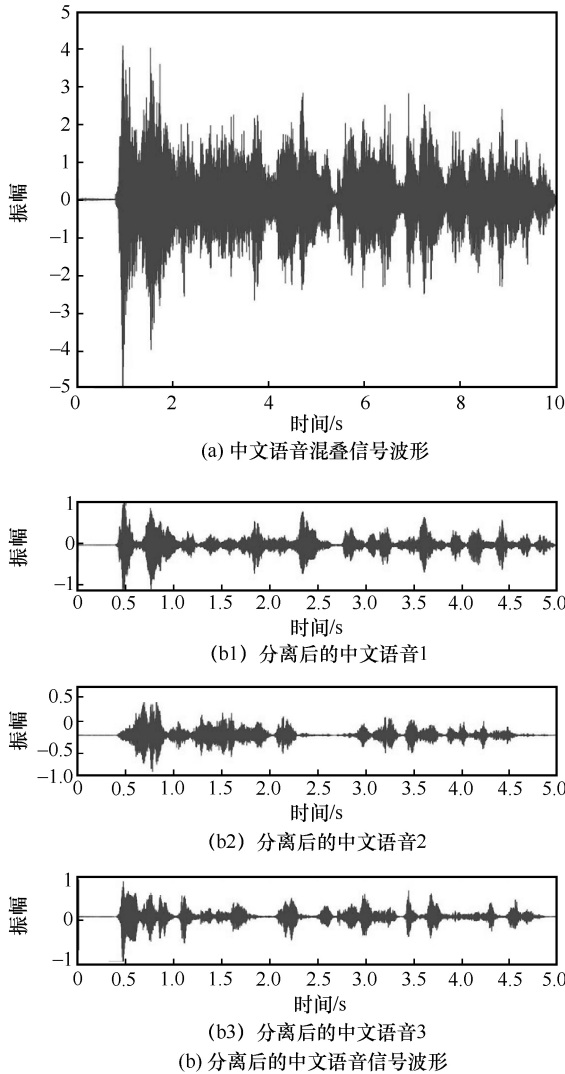


图 8 中文语音混叠信号波形以及分离后的中文语音信号波形

3.6 算法对真实噪声的稳健性分析

为了验证 GIR-UBSS 算法对真实噪声的稳健性，现实生活中遇到的噪声场通常可以近似为球形或圆柱形噪声场，如时间相关噪声、由相互独立的语音片段混合而成的巴布语音或工厂噪声，以及室外测量经常受到各种声源（如交通、自然环境声音等）的干扰。本文实验测试 3 种真实噪声：由球形产生的各向同性噪声 (isotropic noise)、由相互独立的语音片段混合而成的巴布语音噪声 (babble noise) 以及风噪声 (wind noise)。然后分别把这 3 种噪声加入仿真实验 1 中的混叠语音信

号中构成含噪声的混叠信号，利用 GIR-UBSS 算法对混叠信号进行分离，同时对比无噪声 (without noise) 下的盲源分离，实验结果如图 9 所示。从图 9 可知，加入不同噪声以后获得的盲源分离结果与无噪声下得到的结果相似，验证了 GIR-UBSS 算法对真实噪声具有很好的稳健性。

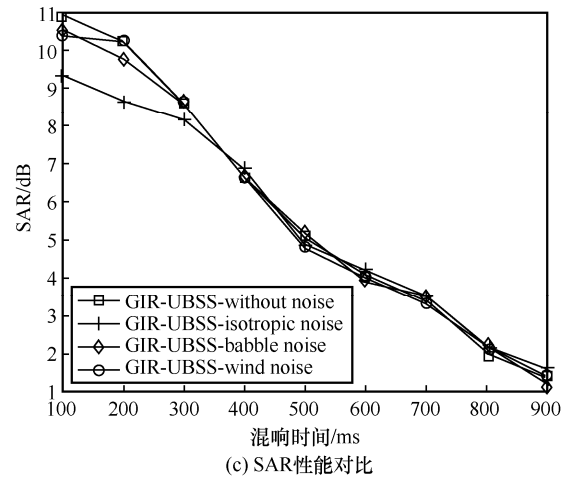
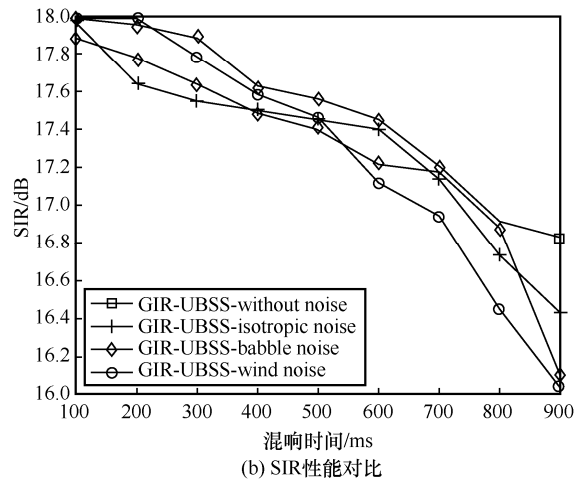
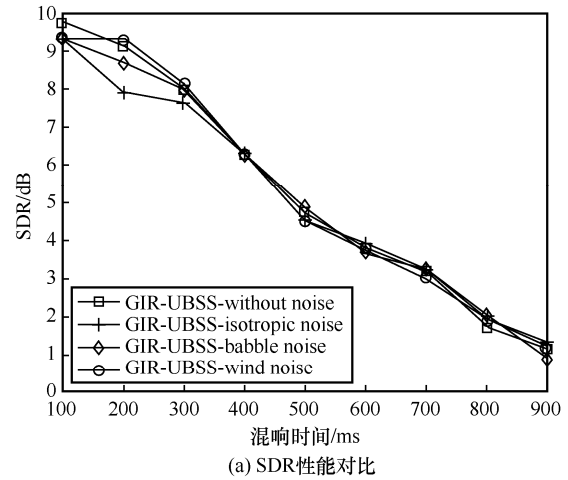


图 9 含有真实噪声的盲源分离性能对比

4 结束语

现实生活中,接收的混叠信号往往伴随高混响等不确定因素,如何消除或削弱高混响的影响,提高盲源分离性能,已经成为信号处理中极具挑战性和现实意义的课题。为此,本文提出一种面向高混响复杂环境的欠定卷积盲源分离算法,通过设计全局脉冲响应网络,减少脉冲响应的长度,削弱可听混响回声的影响。进而构建高混响环境的时频域混叠信号数学模型,设计新模型下的参数更新规则,实现源信号的盲分离。理论分析表明,在新的实时模型更新规则下,可得到频域上的源信号。实验验证了所提 GIR-UBSS 算法对分离中英文语音混叠信号、音乐混叠信号具有很好的有效性。另外,通过与国际上比较流行的盲源分离算法对比,证实了 GIR-UBSS 算法的优越性,以及对真实噪声具有良好的稳健性。

参考文献:

- [1] 张贤达, 保铮. 盲信号分离[J]. 电子学报, 2001, 29(S1): 1766-1771.
ZHANG X D, BAO Z. Blind source separation[J]. Acta Electronica Sinica, 2001, 29(S1): 1766-1771.
- [2] 马昊, 郑翔, 张煜, 等. 盲源分离技术研究: 有监督与无监督[J]. 计算机时代, 2022(11): 59-64.
MA H, ZHENG X, ZHANG Y, et al. Research on blind source separation techniques: supervised vs. unsupervised[J]. Computer Era, 2022(11): 59-64.
- [3] 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.
- [4] 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.
- [5] MIETTINEN J, NITZAN E, VOROBYOV S A, et al. Graph signal processing meets blind source separation[J]. IEEE Transactions on Signal Processing, 2021, 69: 2585-2599.
- [6] 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. Convolutional blind source separation method based on tensor decomposition[J]. Journal on Communications, 2021, 42(8): 52-60.
- [7] 王泽林, 陈镡, 卢晶. 车载场景结合盲源分离与多说话人状态判决的语音抽取[J]. 声学学报, 2020, 45(5): 696-706.
WANG Z L, CHEN K, LU J. Speech extraction based on blind source separation and multi-talker status tracking in automobile environment[J]. Acta Acustica, 2020, 45(5): 696-706.
- [8] 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.
- [9] KRISHNAN L, BETLEHEM T, TEAL P D. Fast algorithms for acoustic impulse response shaping[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(2): 392-403.
- [10] JUNGMANN J O, MAZUR R, KALLINGER M, et al. Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(6): 1829-1842.
- [11] MERTINS A, MAASS M, KATZBERG F. Room impulse response reshaping and crosstalk cancellation using convex optimization[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 489-502.
- [12] 邹亮, 张鹏, 陈勋. 基于三阶统计量的欠定盲源分离方法[J]. 电子与信息学报, 2022, 44(11): 3960-3966.
ZOU L, ZHANG P, CHEN X. Underdetermined blind source separation based on third-order statistics[J]. Journal of Electronics & Information Technology, 2022, 44(11): 3960-3966.
- [13] 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.
- [14] 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, 2022: doi.org/10.16383/j.aas.c211207.
XIE Y, ZOU T, SUN W J, et al. Novel blind source separation method for convolutive mixed environment[J]. Act Autom Sin, 2022: doi.org/10.16383/j.aas.c211207.
- [15] 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.
- [16] 汤辉, 王殊. 基于稳健联合分块对角化的卷积盲源分离[J]. 自动化学报, 2013, 39(9): 1502-1510.
TANG H, WANG S. Robust joint block diagonalization based convolutional blind source separation[J]. Acta Automatica Sinica, 2013, 39(9): 1502-1510.
- [17] ITO N, IKESHITA R, SAWADA H, et al. A joint diagonalization based efficient approach to underdetermined blind audio source separation using the multichannel Wiener filter[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 1950-1965.
- [18] DUONG N Q K, VINCENT E, GRIBONVAL R. Under-determined 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.
- [19] 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.

- [20] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.
- [21] SAWADA H, KAMEOKA H, ARAKI S, et al. Multichannel extensions of non-negative matrix factorization with complex-valued data[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(5): 971-982.
- [22] SEKIGUCHI K, BANDO Y, NUGRAHA A A, et al. Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2610-2625.
- [23] WANG J Y, GUAN S Z, LIU S P, et al. Minimum-volume multichannel nonnegative matrix factorization for blind audio source separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 3089-3103.
- [24] 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.
- [25] WANG T H, YANG F R, YANG J. Convolutional 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] ALLEN J B, BERKLEY D A. Image method for efficiently simulating small-room acoustics[J]. The Journal of the Acoustical Society of America, 1979, 65(4): 943-950.
- [27] 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 (1989-), 男, 安徽利辛人, 博士, 广州大学讲师, 主要研究方向为盲信号分离、信号处理和机器学习等。



邹涛 (1975-), 男, 辽宁沈阳人, 博士, 广州大学教授, 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。



孙为军 (1975-), 男, 安徽马鞍山人, 博士, 广东工业大学副教授, 主要研究方向为模式识别和机器学习等。



谢胜利 (1956-), 男, 湖北荆州人, 博士, 广东工业大学教授, 主要研究方向为无线网络、自动控制、盲信号处理等。