

基于脉冲响应压缩的卷积盲源分离方法

张志伟,高洪元*,王佳仪,刘亚鹏

(哈尔滨工程大学信息与通信工程学院,黑龙江哈尔滨 150001)

摘 要: 已有方法往往将加权预测误差(Weighted Prediction Error, WPE)去混响方法与独立成分分析(Independent Component Analysis, ICA)技术、独立向量分析(Independent Vector Analysis, IVA)技术结合以实现混响环境下的卷积盲源分离。然而,ICA和IVA没有构建灵活的声源模型,因此独立低秩矩阵分析(Independent Low Rank Matrix Analysis, ILRMA)技术被提出,进而组成WPE与ILRMA的联合优化,即WPE-ILRMA。但是,WPE-ILRMA的性能依然依赖迭代次数和滤波器阶数设置,迭代次数较少时,方法性能较差,固定的滤波器阶数设置也无法应对复杂多变的混响环境,并且分离滤波器的更新依然需要对矩阵求逆,消耗硬件资源较大。利用梯度下降法优化预滤波器的脉冲响应重构(Impulse Response Remodeling, IRR)模型被提出以缩短脉冲响应长度和降低高混响环境的影响,但是这些方法基于IRR模型构建的目标函数需要设置范数值,范数值仅凭经验设置,设置不当易导致IRR效果下降,而在盲信号处理场景下,根据IRR效果不断调整目标函数的范数值是不实际的。针对现有卷积盲源分离方法性能依赖于范数值或者滤波器阶数设置的难题,首次将压缩变换概念引入IRR技术中,推导证明了压缩变换可以实现优于 p 范数的混响抑制性能,进而可以用压缩变换替代传统IRR模型中的 p 范数,创新性地设计了一种压缩变换与无穷范数相结合的脉冲响应压缩(Impulse Response Compression, IRC)模型,所提出的IRC模型既实现了更优越的混响抑制性能,又无需像传统去混响方法一样凭经验设置范数值或滤波器阶数以调节混响抑制性能。另外引入迭代源控制(Iterative Source Steering, ISS)技术与ILRMA结合,规避了矩阵求逆运算,使所提方法更易于硬件实现。仿真结果表明,在高混响环境下,不论是针对成分差异较小的动物声音信号分离问题、成分复杂的音乐信号分离问题,还是成分差异较大的语音信号分离问题,所提出的方法相比于其他方法均实现了更优越的信号分离性能,这表明了所提出方法拥有较强的鲁棒性,应用范围广泛。已有方法采用固定的范数值或滤波器阶数设置,参数设置不当会导致方法性能下降,而所提出方法的性能不依赖于范数值或滤波器阶数的调整。所提出的方法突破了已有方法的应用局限,并且IRC技术也可移植应用于其他与去混响相关的工程问题中。

关键词: 去混响;卷积盲源分离;独立低秩矩阵分析;脉冲响应重构;脉冲响应压缩;迭代源控制

基金项目: 国家自然科学基金(No.62372131);黑龙江省自然科学基金(No.LH2020F017);中央高校基本科研业务费专项资金(No.3072025YC0801)

中图分类号: TN912.3

文献标识码: A

文章编号: 0372-2112(2026)06-2805-14

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

DOI: 10.12263/DZXB.20260468

Convolutional Blind Source Separation Method Based on Impulse Response Compression

ZHANG Zhiwei, GAO Hongyuan*, WANG Jiayi, LIU Yapeng

(College of Information and Communication Engineering, Harbin Engineering University, Harbin, Heilongjiang 150001, China)

Abstract: The existing methods often combine the weighted prediction error (WPE) de-reverberation method with independent component analysis (ICA) and independent vector analysis (IVA) techniques to realize convolutional blind source separation in a reverberation environment. However, ICA and IVA do not construct flexible sound source models, so the independent low rank matrix analysis (ILRMA) technique is proposed, which in turn constitutes the joint optimization of WPE and ILRMA, namely WPE-ILRMA. However, the performance of WPE-ILRMA still depends on the number of iterations and the setting of filter order. When the number of iterations is small, the performance of the method is poor, and the fixed filter order setting cannot cope with the complex and changeable reverberation environment. Moreover, the update of the separation filter still needs to invert the matrix, which consumes a lot of hardware resources. The impulse response remodeling (IRR) model using the gradient descent method to optimize the pre-filter is proposed to shorten the impulse response length and reduce the influence of a high reverberation environment. However, the objective function Structured by these methods based on the IRR model needs to set the norm value, which is only set by experience. Improper settings will

easily lead to the decrease of IRR effect. However, it is impractical to continuously adjust the norm value of the objective function according to the IRR effect in a blind signal processing scenario. Aiming at the problem that the performance of existing convolutional blind source separation methods depends on the norm value or filter order setting, the concept of compression transform is introduced into IRR technology for the first time. It is proven that the compression transform can achieve reverberation suppression performance better than p -norm, and then the compression transform can be used to replace p -norm in the traditional IRR model. An innovative impulse response compression (IRC) model combining the compression transform and the infinite norm is designed. The proposed IRC model not only achieves superior reverberation suppression performance, but also does not need to adjust reverberation suppression performance by empirically setting a norm value or filter order like traditional de-reverberation methods. In addition, the iterative source steering (ISS) technology and ILRMA are introduced to avoid the matrix inversion operation, which makes the proposed method easier to implement in hardware. The simulation results show that the proposed method achieves superior signal separation performance compared with other methods in a high reverberation environment, whether for the animal sound signal separation problem with small component difference, the music signal separation problem with complex components, or the speech signal separation problem with large component difference, which shows that the proposed method has strong robustness and a wide application range. The existing methods adopt a fixed norm value or filter order setting. Improper parameter setting will lead to the performance degradation of the method, but the performance of the proposed method does not depend on the adjustment of the norm value or filter order. The proposed method breaks through the application limitations of existing methods, and the IRC technology can also be applied to other engineering problems related to de-reverberation.

Keywords: de-reverberation; convolutional blind source separation; independent low rank matrix analysis; impulse response remodeling; impulse response compression; iterative source steering

Foundation Item(s): National Natural Science Foundation of China (No.62372131); Natural Science Foundation of Heilongjiang Province (No.LH2020F017); Fundamental Research Funds for the Central Universities (No.3072025YC0801)

0 引言

盲源分离 (Blind Source Separation, BSS) 的目标是:在没有任何源信号、噪声和混叠系统的信息作为先验知识的情况下,仅对接收信号进行分析以实现多个未知源的分离。BSS 技术的优势在于不需要传输信道和源信号的信息等先验知识或者仅需要少量先验知识即可实现对信号的处理,传统的依赖先验信息的信号处理方法在没有先验知识的情况下性能会恶化甚至失效,同样先验知识的存在也会降低通信系统的实时性,占用一部分信道带宽,计算量增大。由于 BSS 技术不需要训练序列,并且不要求信源及信道信息等先验知识, BSS 技术具有更大的适用范围和更广的应用前景,并且可以实现高效实时的信号处理^[1]。因此该项技术已经在阵列信号处理、医学信号处理、语音信号处理和机械故障诊断等领域中得到广泛应用^[2-5]。

BSS 方法模型可以分为线性瞬时混合模型和线性卷积混合模型,而卷积混合相比于瞬时混合更加复杂,可以将时域上的卷积混合通过短时傅里叶变换 (Short-Time Fourier Transform, STFT) 转换为频域各个频率点上的瞬时混合,之后在各频点上便可以使用瞬时混合模型下的 BSS 算法进行分离,但是依然需要考虑各频点分离信号的幅度模糊性和排序模糊性对方法性能的影响^[6-8]。另外,传统频域卷积 BSS 方法都

是在线性窄带近似假设下利用 STFT,要求脉冲响应长度远远短于 STFT 窗长度,这在高混响下即脉冲响应长度较长的情况下是不实际的,因此传统卷积 BSS 方法在高混响条件下性能下降^[9]。

虽然现有混响环境下的卷积混合 BSS 方法已取得一定进展,但仍存在亟待解决的工程难题。文献[10]首次提出了一种加权预测误差 (Weighted Prediction Error, WPE) 去混响方法,后与独立成分分析 (Independent Component Analysis, ICA) 技术结合^[11]。之后, WPE 又与独立向量分析 (Independent Vector Analysis, IVA) 技术结合^[12]。然而 ICA 和 IVA 由于没有构建灵活的声源模型,其性能已被证实不如独立低秩矩阵分析 (Independent Low Rank Matrix Analysis, ILRMA) 技术^[13-15],所以文献[16]又提出了一种 WPE 与 ILRMA 联合优化的方法,命名为 WPE-ILRMA,此方法一方面利用 WPE 技术实现混合信号去混响,另一方面利用非负矩阵分解 (Non-negative Matrix Factorization, NMF)^[17]对声源信号的功率谱进行建模。但是,传统的 WPE 技术由于需要引入长预测滤波器对混合信号去混响,导致计算复杂度较高。为了解决此问题,一种低复杂度的 WPE-ILRMA 被提出^[18],即使相比于传统的 WPE-ILRMA 单次迭代时长已经大大减少,但此方法性能依然依赖迭代次数和滤波器阶数设置,迭代次数较少时,方法性能较差,固定的滤波器阶数设置也无法应对复杂多变的混响环境,并且分离滤波器的

更新依然需要对矩阵求逆,消耗硬件资源较大^[19]。文献[20-21]设计了利用梯度下降法优化预滤波器的脉冲响应重构(Impulse Response Remodeling, IRR)模型,缩短脉冲响应长度,降低高混响环境的影响,但是这些方法基于 IRR 模型构建的目标函数需要设置范数值,范数值仅凭经验设置,设置不当易导致 IRR 效果下降,而在盲信号处理场景下,根据 IRR 效果不断调整目标函数的范数值是不实际的。

因此,为了突破现有高混响环境下的卷积 BSS 的性能局限和应用局限,提出一种基于脉冲响应压缩(Impulse Response Compression, IRC)和迭代源控制(Iterative Source Steering, ISS)的卷积 BSS 方法。所提出的卷积 BSS 方法有效提升了高混响环境下的信号分离精度,解决了已有方法性能依赖参数设置的问题。具体贡献如下:

(1)首次将压缩变换概念引入 IRR,并推导证明了压缩变换可以实现优于 p 范数的混响抑制性能,进而可以用压缩变换替代传统 IRR 模型中的 p 范数,创新性地设计了一种压缩变换与无穷范数相结合的 IRC 模型。既实现了更优越的混响抑制性能,又无需像传统去混响方法一样凭经验设置范数值或滤波器阶数以调节混响抑制性能。

(2)所提出方法在包括动物声音、音乐、语音等声源信号以及高混响情况下的性能与已有方法相比优势明显,所提出的 IRC 模型可以应用于解决其他混响环境下的工程问题,并且在 NMF 框架下避免了排序模糊问题。

1 信号模型

1.1 卷积混合信号模型

由于声信号在实际房间内传播时,不可避免地会产生混响和回声,此情况下的信号混合情况复杂,线性瞬时混合模型已无法描述,因此构建基于时间延时 τ 的混叠系统的卷积混合信号模型如下:

$$x_m(t) = \sum_{n=1}^N \sum_{\tau=0}^{L-1} a_{m,n}(\tau) s_n(t-\tau) + b_m(t) \quad (1)$$

其中, $t=1, 2, \dots, T$, t 为采样点数, T 为最大采样点数; $m=1, 2, \dots, M$, M 为接收麦克风数目; $n=1, 2, \dots, N$, N 为声源数目。接收混合信号可以表示为 $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$, 声源信号可以表示为 $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$, 第 n 个声源信号传播到第 m 个麦克风的过程中产生的空间脉冲响应表示为 $\mathbf{a}_{m,n} = [a_{m,n}(0), a_{m,n}(1), \dots, a_{m,n}(L-1)]^T$, L 为脉冲响应长度,信道中加性高斯噪声可以表示为 $\mathbf{b}(t) = [b_1(t), b_2(t), \dots, b_M(t)]^T$ 。 $\mathbf{a}_{m,n}$ 可以构成 L 个 $M \times N$ 维

空间脉冲响应矩阵 $\{\mathbf{A}(0), \mathbf{A}(1), \dots, \mathbf{A}(L-1)\}$, 则卷积混合信号模型也可以表示为

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

1.2 基于全局脉冲响应的卷积混合信号模型

由于在高混响环境下,接收到的混叠信号常常伴随混响回声,极大降低了信号可懂度和质量,因此受传统信道串扰消除和空间脉冲响应重构技术的启发^[20-21],可以设计一种预滤波器与扩音器、接收麦克风共同构建一个信号接收网络,此网络产生的空间脉冲响应称为全局脉冲响应,该网络的物理模型见文献[21],可以在抑制混响回声的同时尽可能减少信息损失。

在接收麦克风前安装 K 个扩音器,第 n 个声源信号传播到第 k 个扩音器的过程中产生的长度为 L_h 的空间脉冲响应表示为 $\mathbf{h}_{k,n} = [h_{k,n}(0), h_{k,n}(1), \dots, h_{k,n}(L_h-1)]^T$ 。文献[20-21]中已表明, $\mathbf{h}_{k,n}$ 必须被等效为可优化的预滤波器,通过不断优化预滤波器以抑制高混响。一方面,若此预滤波器被设置为经验固定,过度抑制混响回声很有可能造成信息损失,抑制不足则会导致方法性能下降,即不能应对复杂多变的混响环境。另一方面,由于在盲信号处理场景下,声源位置往往是未知的,而人为摆放的扩音器和麦克风的位置是已知的,因此需要对全局脉冲响应中的未知变量,即预滤波器 $\mathbf{h}_{k,n}$ 进行寻优。第 k 个扩音器传播到第 m 个麦克风的过程中产生的长度为 L_q 的空间脉冲响应表示为 $\mathbf{q}_{m,k} = [q_{m,k}(0), q_{m,k}(1), \dots, q_{m,k}(L_q-1)]^T$, 则全局脉冲响应产生的长度为 L_g 的空间脉冲响应可以表示为

$$\mathbf{g}_{m,n} = \mathbf{q}_{m,k} * \mathbf{h}_{k,n} = [\mathbf{g}_{m,n}(0), \mathbf{g}_{m,n}(1), \dots, \mathbf{g}_{m,n}(L_g-1)]^T \quad (3)$$

其中, $*$ 代表卷积, 则 $L_g = L_h + L_q - 1$, 所以上式也可表示为

$$\mathbf{g}_{m,n} = \mathbf{Q}_{m,k} \mathbf{h}_{k,n} \quad (4)$$

其中, $\mathbf{Q}_{m,k}$ 表示由 $\mathbf{q}_{m,k}$ 生成的 $L_g \times L_h$ 维卷积矩阵。 $\mathbf{g}_{m,n}$ 可以构成 L_g 个 $M \times N$ 维全局脉冲响应矩阵 $\{\mathbf{G}(0), \mathbf{G}(1), \dots, \mathbf{G}(L_g-1)\}$, 则基于全局脉冲响应的卷积混合信号模型可以表示为

$$\mathbf{x}(t) = \sum_{\tau=0}^{L_g-1} \mathbf{G}(\tau) \mathbf{s}(t-\tau) + \mathbf{b}(t) \quad (5)$$

2 高混响抑制

2.1 窗函数构建

在已有 IRR 技术中需要依赖所选择的窗函数调节全局脉冲响应的能量衰减特性,目标是实现信号信

息保留和高混响抑制的均衡^[20]。为了调节产生的全局脉冲响应,引入两个窗口以构建窗函数,分别为期望窗口 $\bar{\mathbf{w}}_d$ 和不期望窗口 $\bar{\mathbf{w}}_u$,定义如下:

$$\bar{\mathbf{w}}_u = \left[\underbrace{0, 0, \dots, 0}_{L_1+L_2}, \underbrace{\bar{\mathbf{w}}_0^T}_{L_3} \right]^T \quad (6)$$

$$\bar{\mathbf{w}}_d = \left[\underbrace{0, 0, \dots, 0}_{L_1}, \underbrace{1, 1, \dots, 1}_{L_2}, \underbrace{1, 1, \dots, 1}_{L_3} \right]^T \quad (7)$$

$$\bar{\mathbf{w}}_0(l) = \frac{\beta-1}{L_3-1} l + 1 \quad (8)$$

其中, $L_1 = \alpha_1 f_s$, $L_2 = \alpha_2 f_s$, $L_3 = L_g - L_1 - L_2$, $l = 0, 1, \dots, L_3 - 1$, f_s 代表信号采样频率, α_1 、 α_2 和 β 代表窗函数调节因子, $\bar{\mathbf{w}}_0(l)$ 为矢量 $\bar{\mathbf{w}}_0$ 中的第 l 个元素。获得期望脉冲响应 $\mathbf{g}_{m,n}^d = \bar{\mathbf{w}}_d \odot \mathbf{g}_{m,n}$, 不期望脉冲响应 $\mathbf{g}_{m,n}^u = \bar{\mathbf{w}}_u \odot \mathbf{g}_{m,n}$, 其中 $\odot$ 代表 Hadamard 积。

2.2 基于无穷范数和压缩变换的 IRC 技术

已有 IRR 技术往往采用 p 范数模型,通过调节 p 值设计不同范数的算法,实现不同的 IRR 效果,但是在盲信号处理场景和复杂多变的混响环境下实时调节范数值以保证混响抑制效果是不实际的。文献[22]详细分析了在盲信号处理场景下,方法性能不依赖于参数设置的压缩变换技术,因此受文献[22]中压缩变换思想的启发,考虑设计一种无需范数值设置的 IRC 技术应用于卷积 BSS 中^[23]。构建目标函数如下:

$$f(\mathbf{h}_{k,n}) = \ln \left(\frac{f_u(\mathbf{h}_{k,n})}{f_d(\mathbf{h}_{k,n})} \right) \quad (9)$$

其中,目标是找到使目标函数(9)达到最小值的 $\mathbf{h}_{k,n}$, $\ln(\cdot)$ 代表自然对数,并且:

$$f_d(\mathbf{h}_{k,n}) = \max(|\mathbf{g}_{m,n}^d|) \quad (10)$$

$$f_u(\mathbf{h}_{k,n}) = \frac{1}{\sum_{l_g=0}^{L_g-1} \left(\exp(|\mathbf{g}_{m,n}^u(l_g)|) - 1 \right)} \quad (11)$$

其中,期望全局脉冲响应经过无穷范数处理后得到 $f_d(\mathbf{h}_{k,n})$,不期望全局脉冲响应经过所提出的压缩变换处理后得到 $f_u(\mathbf{h}_{k,n})$, $\max(\cdot)$ 代表取最大值。利用梯度下降法求解基于全局脉冲构建的目标函数,以实现声源到扩音器之间产生的脉冲响应的优化,对目标函数求偏导,即

$$\begin{aligned} \nabla_{\mathbf{h}_{k,n}} f(\mathbf{h}_{k,n}) &= \frac{1}{f_u(\mathbf{h}_{k,n})} \nabla_{\mathbf{h}_{k,n}} f_u(\mathbf{h}_{k,n}) \dots \\ &\dots - \frac{1}{f_d(\mathbf{h}_{k,n})} \nabla_{\mathbf{h}_{k,n}} f_d(\mathbf{h}_{k,n}) \end{aligned} \quad (12)$$

其中,

$$\nabla_{\mathbf{h}_{k,n}} f_d(\mathbf{h}_{k,n}) = \text{sign}(\mathbf{g}_{m,n}^d(i_{\text{index}})) \bar{\mathbf{w}}_u(i_{\text{index}}) \mathbf{Q}_{m,k,i_{\text{index}}}^T \quad (13)$$

$$\nabla_{\mathbf{h}_{k,n}} f_u(\mathbf{h}_{k,n}) = - \left(\sum_{l_g=0}^{L_g-1} \left(\exp(|\mathbf{g}_{m,n}^u(l_g)|) - 1 \right) \right)^{-2} \dots \quad (14)$$

$$\dots \bar{\mathbf{Q}}_{m,k}^T \exp(|\mathbf{g}_{m,n}^u|)$$

$$\bar{\mathbf{Q}}_{m,k} = \text{diag}(\text{sign}(\mathbf{g}_{m,n}^u)) \text{diag}(\bar{\mathbf{w}}_d) \mathbf{Q}_{m,k} \quad (15)$$

其中, $\text{diag}(\cdot)$ 代表矩阵对角化, $\text{sign}(\cdot)$ 代表符号函数, i_{index} 代表 $|\mathbf{g}_{m,n}^d|$ 最大值位置索引, $\mathbf{Q}_{m,k,i_{\text{index}}}$ 代表卷积矩阵 $\mathbf{Q}_{m,k}$ 处在位置索引为 i_{index} 的行向量。根据以上分析的更新规则如下:

$$\mathbf{h}_{k,n}^{\varepsilon+1} = \mathbf{h}_{k,n}^{\varepsilon} - \mu \nabla_{\mathbf{h}_{k,n}} f(\mathbf{h}_{k,n}^{\varepsilon}) \quad (16)$$

其中, ε 代表迭代次数, μ 代表学习率,最大迭代次数设置为 ε_{max} 。

2.3 压缩变换推导

所提出的方法不仅要解决已有方法性能对参数设置较为敏感的难题,也应该实现相比于已有方法更优越的性能,接下来从理论角度分析所提方法的优越性。文献[20-21]已经证明 p 范数在设置合适的范数值时可以实现有效的 IRR,而所提出的 IRC 可以在无需参数设置的情况下获得比 p 范数更好的性能,给出如下引理并证明。

引理1 声源到扩音器产生的空间脉冲响应表示为 $\mathbf{h}$,则长度为 L_g 的全局脉冲响应表示为 $\mathbf{g} = \mathbf{Q}\mathbf{h}$, $\mathbf{Q}$ 表示卷积矩阵。由 $\mathbf{h}$ 得到的全局脉冲响应 $\mathbf{g}$ 经过压缩变换得到:

$$f_1(\mathbf{h}) = \frac{1}{\sum_{l_g=0}^{L_g-1} \left(\exp(|\mathbf{g}(l_g)|) - 1 \right)} \quad (17)$$

计算由 $\mathbf{h}$ 得到的全局脉冲响应 $\mathbf{g}$ 的 p 范数为

$$f_2(\mathbf{h}) = \left(\sum_{l_g=0}^{L_g-1} |\mathbf{g}(l_g)|^p \right)^{\frac{1}{p}} \quad (18)$$

其中, p 代表范数值。在高混响环境下,满足 $f_1(\mathbf{h}) < f_2(\mathbf{h})$,即压缩变换可以更好地抑制混响回声。

证明 为了比较 $f_1(\mathbf{h})$ 和 $f_2(\mathbf{h})$ 的大小,需要对 $f_2(\mathbf{h})$ 进行如下变换,利用詹森不等式可得:

$$\frac{1}{\sum_{l_g=0}^{L_g-1} |\mathbf{g}(l_g)|^p} \leq \sum_{l_g=0}^{L_g-1} \frac{\alpha_{l_g}^2}{|\mathbf{g}(l_g)|^p} \quad (19)$$

其中, $\alpha_{l_g} \geq 0$ 代表辅助变量且满足 $\sum_{l_g=0}^{L_g-1} \alpha_{l_g} = 1$ 。当且仅

当 $\alpha_{l_g} = \frac{|\mathbf{g}(l_g)|^p}{\sum_{l_g=0}^{L_g-1} |\mathbf{g}(l_g)|^p}$ 时,等号成立。若令 $\alpha_{l_g} = 1/L_g$,则式(19)变为

$$\frac{1}{\sum_{l_g=0}^{L_g-1} |g(l_g)|^p} \leq \sum_{l_g=0}^{L_g-1} \frac{1}{L_g^2 |g(l_g)|^p} = \sum_{l_g=0}^{L_g-1} \frac{1}{L_g^2} |g(l_g)|^{-p} \quad (20)$$

为了将式(20)与 $f_1(\mathbf{h})$ 建立联系,令 $u = |g(l_g)| \geq 0$, 构建如下函数:

$$H(u) = \exp(u) - u^{-p} \quad (21)$$

对函数(21)求导,得:

$$H'(u) = \exp(u) + pu^{-p-1} \quad (22)$$

则注意到 $H(0) = H'(0) = 1 > 0$,进一步证明函数 $H(u)$ 在 $(0, +\infty)$ 上为单调递增函数。 $\forall k_1, k_2 \in (0, +\infty)$ 且 $k_1 < k_2$, 并且由文献[20-21]可知一般设置范数值 $10 \leq p \leq 20$,进而可以得到:

$$H(k_1) - H(k_2) = \exp(k_1) - \exp(k_2) - k_1^{-p} + k_2^{-p} \\ = (\exp(k_1) - \exp(k_2)) + (k_2^{-p} - k_1^{-p}) \quad (23)$$

可知 $H(k_1) - H(k_2) < 0$, $H(k_1) < H(k_2)$, 即函数 $H(u)$ 在 $(0, +\infty)$ 上单调递增。又因为 $H'(0) = 1 > 0$, 因此以下陈述成立: $\exists \delta > 0$, 对于任意 $u \in (0, \delta)$, $H(u) > H(0) = 1 > 0$, 即 $\exp(u) - 1 > u^{-p}$, 所以可得:

$$\sum_{l_g=0}^{L_g-1} \frac{1}{L_g^2} \left(\exp(|g(l_g)|) - 1 \right) > \sum_{l_g=0}^{L_g-1} \frac{1}{L_g^2} |g(l_g)|^{-p} \geq \frac{1}{\sum_{l_g=0}^{L_g-1} |g(l_g)|^p} \quad (24)$$

进一步地:

$$\frac{1}{\sum_{l_g=0}^{L_g-1} \left(\exp(|g(l_g)|) - 1 \right)} < \frac{1}{\sum_{l_g=0}^{L_g-1} |g(l_g)|^{-p}} \leq \frac{1}{L_g^2} \sum_{l_g=0}^{L_g-1} |g(l_g)|^p \quad (25)$$

又因为在文献[20-21]中, $L_g = L_h + L_q - 1 = 2f_s R_T - 1$, 其中 f_s 代表信号采样频率, R_T 代表混响时间, 由此可知 L_g 一般为大于 10^2 的整数, L_g^2 则一般大于 10^4 , 高混响环境下即 $R_T > 500$ ms时, L_g^2 的值会非常大, 而 $10 \leq p \leq 20$, 所以在高混响环境下可确定:

$$\frac{1}{\sum_{l_g=0}^{L_g-1} \left(\exp(|g(l_g)|) - 1 \right)} < \frac{1}{L_g^2} \sum_{l_g=0}^{L_g-1} |g(l_g)|^p < \left(\sum_{l_g=0}^{L_g-1} |g(l_g)|^p \right)^{\frac{1}{p}} \quad (26)$$

综上所述, 在高混响环境下, $f_1(\mathbf{h}) < f_2(\mathbf{h})$ 得证。证毕。

2.4 计算复杂度分析

使用梯度下降法求解目标函数(9)。根据2.1节,

每次迭代需要计算全局脉冲响应的期望脉冲响应和不期望脉冲响应, 产生的计算复杂度为

$$O_1 = O(L_g L_h + 2L_g) \quad (27)$$

在式(12)中, 结合式(13)~式(15)分析可知, 每次迭代计算 $\left(1/f_u(\mathbf{h}_{k,n})\right) \nabla_{\mathbf{h}_{k,n}} f_u(\mathbf{h}_{k,n})$ 产生的计算复杂度为 $O(L_g + L_h + 3)$, 每次迭代计算 $\left(1/f_d(\mathbf{h}_{k,n})\right) \nabla_{\mathbf{h}_{k,n}} f_d(\mathbf{h}_{k,n})$ 产生的计算复杂度为 $O(L_g^2 L_h + L_g L_h + L_g + L_h + 1)$, 所以每次迭代整个式(12)产生的计算复杂度为

$$O_2 = O(L_g^2 L_h + L_g L_h + 2L_g + 3L_h + 4) \quad (28)$$

另外, 式(16)每次迭代产生的计算复杂度为 $O_3 = O(2L_h)$ 。因此, 当最大迭代次数为 $\varepsilon_{\max}$ 时, IRC技术产生的总计算复杂度为

$$O_{\text{IRC}} = \varepsilon_{\max} (O_1 + O_2 + O_3) \quad (29)$$

3 卷积BSS算法

3.1 NMF模型

由于仅在时域上处理卷积混合信号的算法依赖于ICA技术^[11], 算法复杂度高、计算量大且难以硬件实现, 因此转换至频域对信号进行处理的卷积BSS算法逐渐成为研究热点, 而NMF模型可以满足声源信号功率谱建模, 因此被广泛应用于频域卷积BSS算法中。基于全局脉冲响应的卷积混合信号经过STFT可以表示为

$$\mathbf{X}(\bar{f}, \bar{t}) = \mathbf{G}_{\bar{f}} \mathbf{S}(\bar{f}, \bar{t}) + \mathbf{B}(\bar{f}, \bar{t}) \quad (30)$$

其中, $\mathbf{G}_{\bar{f}}$ 代表在频点 $\bar{f}$ 下的混合矩阵, $\bar{f} = 1, 2, \dots, \bar{F}$ 和 $\bar{t} = 1, 2, \dots, \bar{T}$ 分别代表频点和时间帧, $\bar{F}$ 和 $\bar{T}$ 分别代表最大频点和最大时间帧。源信号 $\mathbf{S}_n$ 的 $\bar{F} \times \bar{T}$ 维方差矩阵 $\mathbf{R}_n$ 满足非负矩阵分解, $n = 1, 2, \dots, N$, 在此模型中假设源信号符合零均值且独立同分布, 所以 $\mathbf{R}_n$ 中的 $\bar{f}$ 行 $\bar{t}$ 列元素 $r_{\bar{f}\bar{t}}^n$ 为 $|S_n(\bar{f}, \bar{t})|^2$ 的期望值。NMF模型构建如下:

$$\mathbf{R}_n = \mathbf{Z}_n \mathbf{V}_n \quad (31)$$

其中, $\mathbf{Z}_n$ 和 $\mathbf{V}_n$ 分别代表 $\bar{F} \times J$ 维基矩阵和 $J \times \bar{T}$ 维激活矩阵, J 代表NMF模型的基数。定义频点 $\bar{f}$ 下的 $N \times M$ 维解混矩阵为 $\mathbf{W}_{\bar{f}}$, 则恢复频域源信号为

$$\mathbf{Y}(\bar{f}, \bar{t}) = \mathbf{W}_{\bar{f}} \mathbf{X}(\bar{f}, \bar{t}) \quad (32)$$

研究基于传统ILRMA框架下的卷积BSS方法, 因此假设麦克风数目等于源信号数目, 即 $N = M$ ^[13-15]。通过最小化负对数似然函数可以求解 $\mathbf{W}_{\bar{f}}$, 非负矩阵 $\mathbf{Z}_n$ 和 $\mathbf{V}_n$, 负对数似然函数表示如下:

$$\mathcal{L} = -2\bar{T} \sum_{\bar{f}=1}^{\bar{F}} \left| \det(\mathbf{W}_{\bar{f}}) \right| \dots$$

$$\dots + \sum_{\bar{f}=1}^{\bar{F}} \sum_{\bar{t}=1}^{\bar{T}} \sum_{n=1}^N \left(\frac{\left| \mathbf{w}_{\bar{f},n}^H \mathbf{X}(\bar{f}, \bar{t}) \right|^2}{\sum_{j=1}^J z_{\bar{f},j}^n v_{j,\bar{t}}^n} + \log \sum_{j=1}^J z_{\bar{f},j}^n v_{j,\bar{t}}^n \right) \quad (33)$$

其中, $\mathbf{w}_{\bar{f},n}$ 为解混矩阵 $\mathbf{W}_{\bar{f}}$ 的列向量, $z_{\bar{f},j}^n$ 和 $v_{j,\bar{t}}^n$ 分别代表非负矩阵 $\mathbf{Z}_n$ 的 $\bar{f}$ 行 j 列元素和非负矩阵 $\mathbf{V}_n$ 的 j 行 $\bar{t}$ 列元素, $\det(\cdot)$ 代表取矩阵的行列式。

3.2 基于 ISS 的 ILRMA 算法

在针对声信号卷积 BSS 的工程问题上, 基于 NMF 模型的 ILRMA 的性能已被证实优于 ICA 和 IVA, 而传统 ILRMA 算法每次迭代都需要计算每个频点的协方差矩阵以及矩阵的逆, 会导致较大的硬件资源消耗^[19], 基于 ISS 的 ILRMA 算法则解决了这一工程难题, 为使所提方法更贴合硬件实现与实际应用, 所提方法采用基于 ISS 的 ILRMA 算法实现卷积 BSS。

初始化非负矩阵 $\mathbf{Z}_n^\gamma$ 和 $\mathbf{V}_n^\gamma$, 矩阵中元素均为在 $[0, 1]$ 间服从均匀分布的随机数, γ 代表迭代次数, 最大迭代次数设置为 $\gamma_{\max}$, $(z_{\bar{f},j}^n)^\gamma$ 和 $(v_{j,\bar{t}}^n)^\gamma$ 分别代表非负矩阵 $\mathbf{Z}_n^\gamma$ 的 $\bar{f}$ 行 j 列元素和非负矩阵 $\mathbf{V}_n^\gamma$ 的 j 行 $\bar{t}$ 列元素。初始化频点 $\bar{f}$ 下解混矩阵 $\mathbf{W}_{\bar{f}}$ 为单位矩阵, 在频点 $\bar{f}$ 上的 $N \times \bar{T}$ 维声源信号矩阵表示为 $\mathbf{Y}_{\bar{f}}^\gamma$, $(y_{n,\bar{t}}^\gamma)^\gamma$ 代表 $\mathbf{Y}_{\bar{f}}^\gamma$ 的 n 行 $\bar{t}$ 列元素。 $\mathbf{R}_n^\gamma = \mathbf{Z}_n^\gamma \mathbf{V}_n^\gamma$, $(r_{\bar{f},\bar{t}}^n)^\gamma$ 代表矩阵 $\mathbf{R}_n^\gamma$ 的 $\bar{f}$ 行 $\bar{t}$ 列元素。首先, 负对数似然函数的最小化可以通过迭代以下空间模型参数的更新规则来执行:

$$(z_{\bar{f},j}^n)^{\gamma+1} = (z_{\bar{f},j}^n)^\gamma \frac{\sum_{\bar{t}=1}^{\bar{T}} \left| (y_{n,\bar{t}}^\gamma)^\gamma \right|^2 \left((r_{\bar{f},\bar{t}}^n)^\gamma \right)^{-2} (v_{j,\bar{t}}^n)^\gamma}{\sum_{\bar{t}=1}^{\bar{T}} \left((r_{\bar{f},\bar{t}}^n)^\gamma \right) (v_{j,\bar{t}}^n)^\gamma} \quad (34)$$

$$(\bar{r}_{\bar{f},\bar{t}}^n)^\gamma = \sum_{j=1}^J (z_{\bar{f},j}^n)^{\gamma+1} (v_{j,\bar{t}}^n)^\gamma \quad (35)$$

$$(v_{j,\bar{t}}^n)^{\gamma+1} = (v_{j,\bar{t}}^n)^\gamma \frac{\sum_{\bar{f}=1}^{\bar{F}} \left| (y_{n,\bar{t}}^\gamma)^\gamma \right|^2 \left((\bar{r}_{\bar{f},\bar{t}}^n)^\gamma \right)^{-2} (z_{\bar{f},j}^n)^{\gamma+1}}{\sum_{\bar{f}=1}^{\bar{F}} \left((\bar{r}_{\bar{f},\bar{t}}^n)^\gamma \right) (z_{\bar{f},j}^n)^{\gamma+1}} \quad (36)$$

$$(r_{\bar{f},\bar{t}}^n)^{\gamma+1} = \sum_{j=1}^J (z_{\bar{f},j}^n)^{\gamma+1} (v_{j,\bar{t}}^n)^{\gamma+1} \quad (37)$$

引入 ISS 矢量对声源信号进行更新如下:

$$\mathbf{Y}_{\bar{f}}^{\gamma+1} = \mathbf{Y}_{\bar{f}}^\gamma - \bar{\mathbf{v}}_{\bar{f},n}^\gamma \mathbf{Y}_{\bar{f},n}^\gamma \quad (38)$$

其中: $\mathbf{Y}_{\bar{f},n}^\gamma$ 代表更新前频点 $\bar{f}$ 上的第 n 路 $1 \times \bar{T}$ 维声源信号; $\bar{\mathbf{v}}_{\bar{f}}^\gamma$ 代表频点 $\bar{f}$ 上的 $N \times 1$ 维 ISS 矢量, 对矢量 $\bar{\mathbf{v}}_{\bar{f}}^\gamma$ 中元

素 $\bar{v}_{\bar{f},n}^\gamma$ 进行更新, $\bar{n} = 1, 2, \dots, N$ 。文献[24]推导了 ISS 的更新规则如下。当 $\bar{n}$ 与 n 不相等时:

$$\bar{v}_{\bar{f},\bar{n}}^\gamma = \frac{u_{\bar{f},\bar{n}}^\gamma}{d_{\bar{f},\bar{n}}^\gamma} \quad (39)$$

$$d_{\bar{f},\bar{n}}^\gamma = \frac{1}{\bar{T}} \sum_{\bar{t}=1}^{\bar{T}} \frac{\left| (y_{n,\bar{t}}^\gamma)^\gamma \right|^2}{(r_{\bar{f},\bar{t}}^n)^{\gamma+1}} \quad (40)$$

$$u_{\bar{f},\bar{n}}^\gamma = \frac{1}{\bar{T}} \sum_{\bar{t}=1}^{\bar{T}} \frac{(y_{n,\bar{t}}^\gamma)^\gamma \text{conj}\left((y_{n,\bar{t}}^\gamma)^\gamma\right)}{(r_{\bar{f},\bar{t}}^n)^{\gamma+1}} \quad (41)$$

当 $\bar{n}$ 与 n 相等时:

$$\bar{v}_{\bar{f},\bar{n}}^\gamma = 1 - \frac{1}{\sqrt{d_{\bar{f},\bar{n}}^\gamma}} \quad (42)$$

最后, 对更新后的源基矩阵和声源信号进行归一化。 $(y_{n,\bar{t}}^\gamma)^{\gamma+1}$ 代表频点 $\bar{f}$ 上的声源信号矩阵 $\mathbf{Y}_{\bar{f}}^{\gamma+1}$ 的 n 行 $\bar{t}$ 列元素, 则归一化操作如下:

$$(y_{n,\bar{t}}^\gamma)^{\gamma+1} \leftarrow \frac{(y_{n,\bar{t}}^\gamma)^{\gamma+1}}{\sqrt{\frac{1}{\bar{F}\bar{T}} \sum_{\bar{f}=1}^{\bar{F}} \sum_{\bar{t}=1}^{\bar{T}} \left| (y_{n,\bar{t}}^\gamma)^{\gamma+1} \right|^2}} \quad (43)$$

$$(r_{\bar{f},\bar{t}}^n)^{\gamma+1} \leftarrow \frac{(r_{\bar{f},\bar{t}}^n)^{\gamma+1}}{\frac{1}{\bar{F}\bar{T}} \sum_{\bar{f}=1}^{\bar{F}} \sum_{\bar{t}=1}^{\bar{T}} \left| (y_{n,\bar{t}}^\gamma)^{\gamma+1} \right|^2} \quad (44)$$

$$(z_{\bar{f},j}^n)^{\gamma+1} \leftarrow \frac{(z_{\bar{f},j}^n)^{\gamma+1}}{\frac{1}{\bar{F}\bar{T}} \sum_{\bar{f}=1}^{\bar{F}} \sum_{\bar{t}=1}^{\bar{T}} \left| (y_{n,\bar{t}}^\gamma)^{\gamma+1} \right|^2} \quad (45)$$

3.3 反向投影法消除各频点排序模糊

频域卷积 BSS 算法还需要考虑各频点分离信号的模糊性对方法性能的影响^[6-8], 因此引入文献[14]中的反向投影法消除各频点排序模糊性。选择在频点 $\bar{f}$ 上的第一路 $1 \times \bar{T}$ 维卷积混合信号 $\mathbf{X}_{\bar{f},1}$ 作为参考信号, $\mathbf{Y}_{\bar{f}}$ 表示 NMF 迭代模型迭代结束后输出的频点 $\bar{f}$ 上的 $N \times \bar{T}$ 维声源分离信号。反向投影矢量计算如下:

$$\boldsymbol{\theta}_{\bar{f}} = \frac{\mathbf{X}_{\bar{f},1} \mathbf{Y}_{\bar{f}}^H}{\mathbf{Y}_{\bar{f}} \mathbf{Y}_{\bar{f}}^H} \quad (46)$$

其中, $\boldsymbol{\theta}_{\bar{f}}$ 代表在频点 $\bar{f}$ 上的 $1 \times N$ 维反向投影矢量, $\boldsymbol{\theta}_{\bar{f}}$ 中的第 n 维元素可以表示为 $\theta_{\bar{f},n}$, 则得到如下反向投影公式:

$$\hat{\mathbf{Y}}_{\bar{f},n} = \theta_{\bar{f},n} \mathbf{Y}_{\bar{f},n} \quad (47)$$

其中: $\mathbf{Y}_{\bar{f},n}$ 代表 $\mathbf{Y}_{\bar{f}}$ 的第 n 个 $1 \times \bar{T}$ 维声源分离信号; $\hat{\mathbf{Y}}_{\bar{f},n}$ 则代表反向投影后的信号, 所有频点的各路分离信号组成 $\hat{\mathbf{Y}}$, 对 $\hat{\mathbf{Y}}$ 进行短时傅里叶逆变换即可得到最终的

时域声源分离信号 $\hat{\mathbf{y}}$ 。

所提出方法命名为 IRC-ILRMAISS, 其算法流程图如图 1 所示, IRC 算法伪代码如算法 1 所示, ILRMAISS 算法伪代码如算法 2 所示。

算法 1 IRC

输入: 初始全局脉冲响应和窗函数

输出: 抑制混响后的全局脉冲响应

FOR $k=1, 2, \dots, K$

FOR $n=1, 2, \dots, N$

设置迭代次数 $\varepsilon=1$;

WHILE $\varepsilon \leq \varepsilon_{\max}$

通过式(12)~(15)求目标函数(9)的

偏导;

通过式(16)更新声源到扩音器之间产生的脉冲响应;

设置 $\varepsilon = \varepsilon + 1$;

END WHILE

通过式(4)计算声源到麦克风产生的全局脉冲响应;

END FOR

END FOR

算法 2 ILRMAISS

输入: 基于全局脉冲响应的时域卷积混合信号 $\mathbf{x}$

输出: 时域声源分离信号 $\hat{\mathbf{y}}$

$\mathbf{x}$ 经过短时傅里叶变换得到频域卷积混合信号 $\mathbf{X}$;

初始化非负矩阵和解混矩阵, 基于 $\mathbf{X}$ 得到初始化的声源信号;

设置迭代次数 $\gamma=1$;

WHILE $\gamma \leq \gamma_{\max}$

通过式(34)~(37)更新非负矩阵;

通过式(38)~(42)更新声源信号;

通过式(43)~(45)对更新后的非负矩阵和声源信号进行归一化;

设置 $\gamma = \gamma + 1$;

END WHILE

通过式(46)~(47)消除各频点排序模糊性, 得到频域声源分离信号 $\hat{\mathbf{Y}}$;

$\hat{\mathbf{Y}}$ 经过逆短时傅里叶变换得到时域声源分离信号 $\hat{\mathbf{y}}$;

3.4 计算复杂度分析

根据 3.2 节, 式(34)每次迭代产生的计算复杂度为 $O(N(3\bar{F}J + 2\bar{F}\bar{T}(1+J)))$, 式(35)每次迭代产生的计算复杂度为 $O(N(\bar{F}J\bar{T}))$, 式(36)每次迭代产生的计算复杂度为 $O(N(3J\bar{T} + 2\bar{F}\bar{T}(1+J)))$, 式(37)每次迭代产生的计算复杂度为 $O(N(\bar{F}J\bar{T}))$, 综上所述, 更新空间模型参数每次迭代产生的计算复杂度为

$$O_1 = O(N(3J(\bar{F} + \bar{T}) + \bar{F}\bar{T}(4+6J))) \quad (48)$$

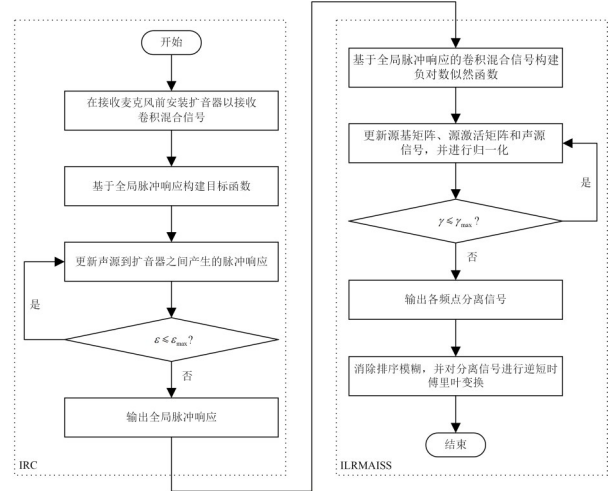


图 1 IRC-ILRMAISS 算法流程图

Figure 1 Algorithm flow chart of IRC-ILRMAISS

当 $\bar{n}$ 与 n 不相等时, 式(39)每次迭代产生的计算复杂度为 $O(N(N-1)\bar{F}(4\bar{T}+1))$, 当 $\bar{n}$ 与 n 相等时, 式(42)每次迭代产生的计算复杂度为 $O(N\bar{F}(2\bar{T}+3))$, 式(38)每次迭代产生的计算复杂度为 $O(2N^2\bar{F}\bar{T})$ 。综上所述, 引入迭代源控制矢量对声源信号进行更新, 每次迭代产生的计算复杂度为

$$O_2 = O(N\bar{F}(6N\bar{T} + N - 2\bar{T} + 2)) \quad (49)$$

最后对更新后的源基矩阵和声源信号进行归一化, 式(43)~式(45)每次迭代产生的计算复杂度为

$$O_3 = O(3\bar{F}\bar{T}N + \bar{F}JN + N) \quad (50)$$

因此, 当最大迭代次数为 $\gamma_{\max}$ 时, ILRMAISS 算法产生的总计算复杂度为

$$O_{\text{ILRMAISS}} = \gamma_{\max}(O_1 + O_2 + O_3) \quad (51)$$

4 仿真实验

为了验证所提出的卷积 BSS 方法的有效性, 引入广泛应用的三种 BSS 性能评价指标^[25], 分别为信号干扰比 (Source-to-Interference Ratio, SIR)、信号失真比 (Source-to-Distortion Ratio, SDR) 和信号伪像比 (Source-to-Artifacts Ratio, SAR), 其值越大, BSS 性能越好。将所提出的方法记为 IRC-ILRMAISS, 与传统 ILRMA 算法^[13]、一致 ILRMA 算法即 Consistent-ILRMA^[14]、WPE-ILRMA^[18]以及文献[20]中的 IRR 方法分别与传统 ILRMA 算法和 ILRMAISS 算法结合产生的 IRR-ILRMA 和 IRR-ILRMAISS 算法进行比较。根据文献[18], WPE-ILRMA 的滤波器长度设置为 5。根据文献[20], IRR-ILRMA 和 IRR-ILRMAISS 的 p 范数设置为 10。为保证实验公平性, 所有方法 NMF 模型基数和算法最大迭代次数均保持一致, 对比方法的其

他参数设置参考各自参考文献。

利用文献[26]中的环境模拟方法,也是国际公用的方法,构建一个混响空间,空间大小为 $1\text{ m} \times 1\text{ m} \times 2\text{ m}$,放置两个声源位置为 $[0.5\ 0\ 0.36]$ 和 $[0.5\ 0.3\ 0.36]$,两个扩音器位置为 $[0.6\ 0.05\ 0.26]$ 和 $[0.6\ 0.25\ 0.26]$,两个麦克风位置为 $[0.7\ 0.1\ 0.16]$ 和 $[0.7\ 0.2\ 0.16]$ 。参考文献[20]中的模型参数,本方法参数设置如下: $M=N=K=2$,采样频率 $f_s=11.025\text{ kHz}$, $L_q=L_h=0.05f_sR_T$, $\alpha_1=\alpha_2=0.004$, $\beta=2$, $\mu=3 \times 10^{-3}$, $\varepsilon_{\max}=300$, $\gamma_{\max}=50$, $J=10$,信号采样时长均为 1 s 。对于STFT,汉明窗口的长度被设置为512,跳跃大小被设置为20,快速傅里叶变换点的数量被设置为512。仿真实验均在MATLABR2024a上实现,硬件信息如下: Intel(R)Core(TM)i5-10300H CPU

(2.50 GHz),在微软Windows 11上处理16 GB的RAM。

4.1 动物声音信号实验

选择文献[26]中的动物声音信号“hengjiao”和“paoxiao”作为声源信号,在混响时间为700 ms的情况下测试所提出的IRC-ILRMAISS的方法性能,源信号、卷积混合信号以及分离信号的时域波形分别如图2所示。在不考虑噪声影响的情况下,混响时间从400 ms增加至700 ms,每个混响时间下进行50次仿真实验,取平均值作为实验结果,仿真结果如图3所示。在考虑混响时间为700 ms的情况下,信噪比从10 dB增加至30 dB,每个信噪比(Signal-to-Noise Ratio, SNR)下进行50次仿真实验,取平均值作为实验结果,仿真结果如图4所示。

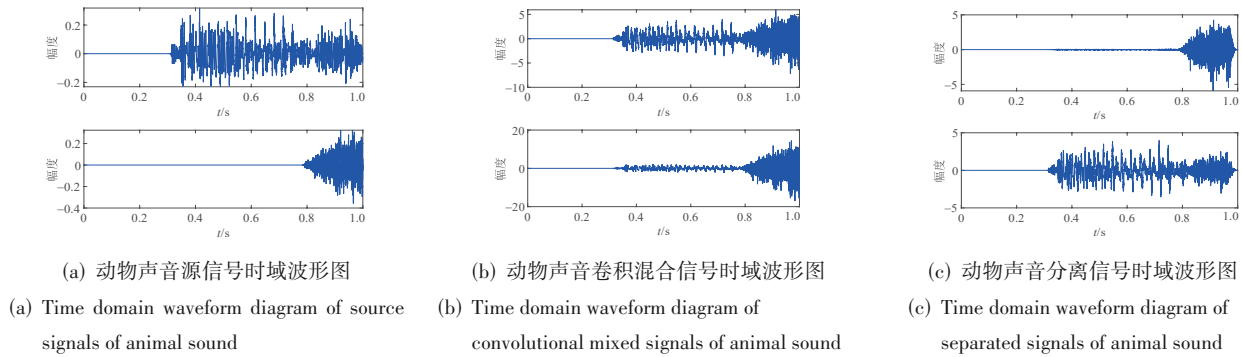


图2 高混响下基于动物声音信号的IRC-ILRMAISS盲源分离效果图

Figure 2 Blind source separation effect diagram of IRC-ILRMAISS based on animal sound signals under high reverberation

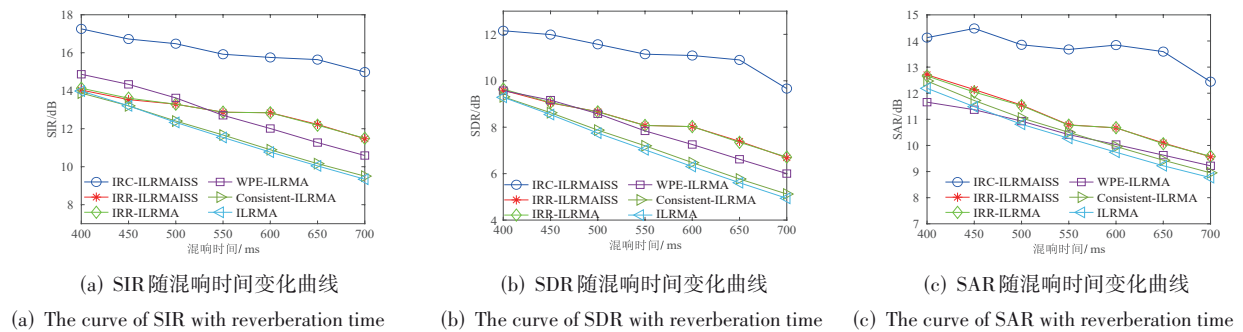


图3 基于动物声音信号的盲源分离性能与混响时间的关系

Figure 3 The relationship between blind source separation performance and reverberation time based on animal sound signals

由图2可知,所提出的IRC-ILRMAISS可以在高混响情况下实现有效的声源信号分离,此时得到的SIR、SDR和SAR分别为16.324 2 dB、11.296 2 dB和13.700 6 dB。图3(a)~(c)分别为针对动物声音信号分离的SIR、SDR和SAR随混响时间变化曲线,BSS性能随着混响时间的增长不断下降。图4(a)~(c)分别为针对动物声音信号分离的SIR、SDR和SAR随SNR变化的曲

线,BSS性能随着SNR的增长不断提升。由图3(a)~(b)可以看出,IRR-ILRMA、IRR-ILRMAISS和WPE-ILRMA相比于没有抑制混响的ILRMA和Consistent-ILRMA性能更优,但在图3(c)中WPE-ILRMA与Consistent-ILRMA性能近似。SAR反映分离算法对噪声的抑制能力^[25],WPE-ILRMA在SIR和SDR上性能良好但在SAR上较差,说明WPE-ILRMA抗噪能力不足。所以在

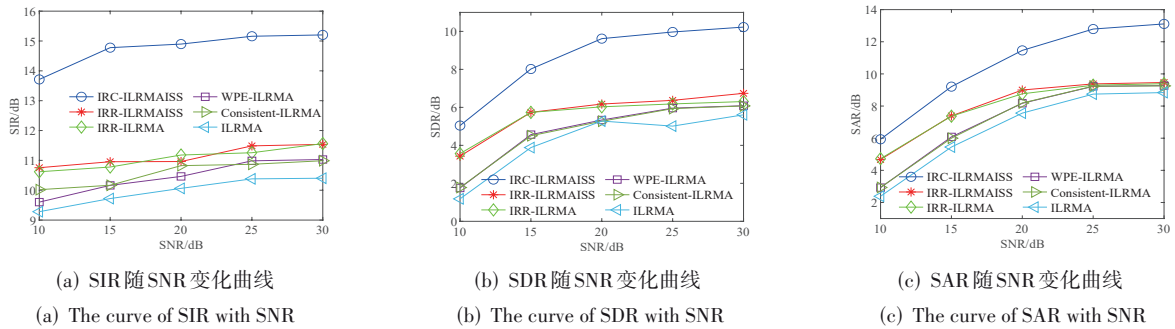


图4 基于动物声音信号的盲源分离性能与SNR的关系

Figure 4 The relationship between blind source separation performance and SNR based on animal sound signals

图4(a)~(c)中,WPE-ILRMA在三种指标上与没有抑制混响的 Consistent-ILRMA 性能几乎没有区别,也说明 WPE-ILRMA 在噪声影响下混响抑制性能下降。反观所提出的 IRC-ILRMAISS,一方面保留了 IRR-ILRMA 优秀的抗噪性能,另一方面进一步提升了 IRR-ILRMA 的混响抑制效果,相比于其他方法取得了显著的分性性能,实验进一步证实了 2.3 节中的引理,即压缩变换可以实现相较于 p 范数更优的 IRR 效果,进而可以提升高混响环境下的 BSS 精度。IRR-ILRMA 和 IRR-ILRMAISS 的性能无明显区别,此结果与文献[24]中的实验结果分析一致,说明方法性能的大幅提升并不是由 ISS 引起的,所提出的 IRC 确实有效抑制了高混响。

所提出的方法能取得优势的原因在于:在高混响环境下,IRR 和 WPE 的混响抑制性能分别受固定的 p 范数设置和固定的滤波器阶数设置的限制,导致方法性能的提升有限。对 p 范数和滤波器阶数的设置进行进一步调节,确实有可能进一步提升 BSS 性能,但是这在没有任何先验知识的盲信号处理场景下是不实际的。所设计的 IRC 模型无需设置范数值或者滤波器阶数,IRC 效果不受参数设置的影响,更适用于

盲信号处理场景,也可以移植应用于其他与混响抑制相关的工程问题中。

4.2 音乐信号实验

选择文献[13]中的音乐信号“piano”和“drums”作为声源信号,其他参数设置与 4.1 节保持一致。仿真结果如图 5、图 6 和图 7 所示。由图 5(a)~(c)可知,两路音乐信号在高混响情况下依然被有效分离,此时得到的 SIR、SDR 和 SAR 分别为 15.046 1 dB、10.871 8 dB 和 14.532 1 dB。图 6(a)~(c)分别为针对音乐信号分离的 SIR、SDR 和 SAR 随混响时间变化曲线。图 7(a)~(c)分别为针对音乐信号分离的 SIR、SDR 和 SAR 随 SNR 变化曲线。由仿真结果图 6 和图 7 可知,针对高混响环境下音乐信号的分离问题,WPE-ILRMA 的性能恶化甚至不如未采取混响抑制的方法。分析其原因可能是相比于动物声音信号,音乐信号的源成分更加复杂^[21],WPE-ILRMA 目前的滤波器阶数设置不能应对在更加复杂的信号成分下的混响抑制,甚至可能造成信号的信息损失。所提出的 IRC-ILRMAISS 则保留了 IRR-ILRMA 在成分复杂的音乐信号分离情况下的混响抑制能力,并且相比于其他方法性能更优。

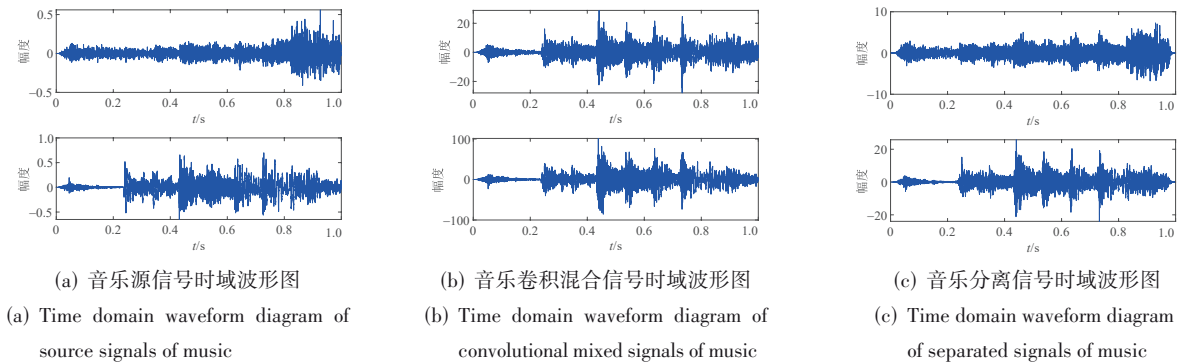


图5 高混响下基于音乐信号的 IRC-ILRMAISS 盲源分离效果图

Figure 5 Blind source separation effect diagram of IRC-ILRMAISS based on music signals under high reverberation

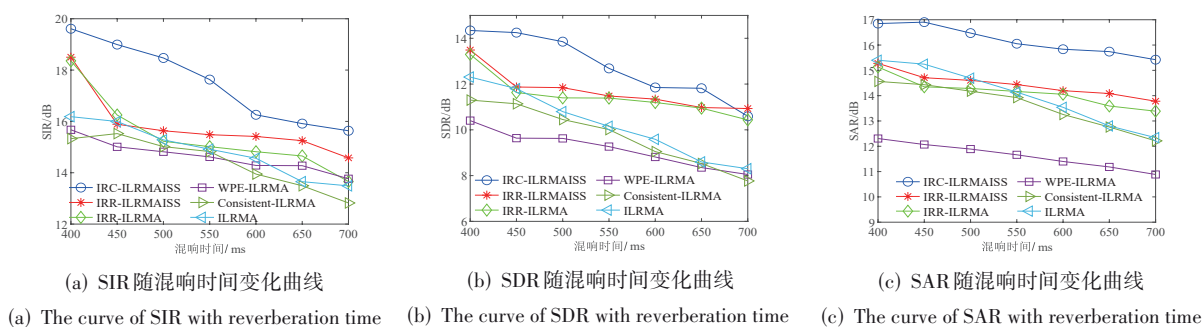


图6 基于音乐信号的盲源分离性能与混响时间的关系

Figure 6 The relationship between blind source separation performance and reverberation time based on music signals

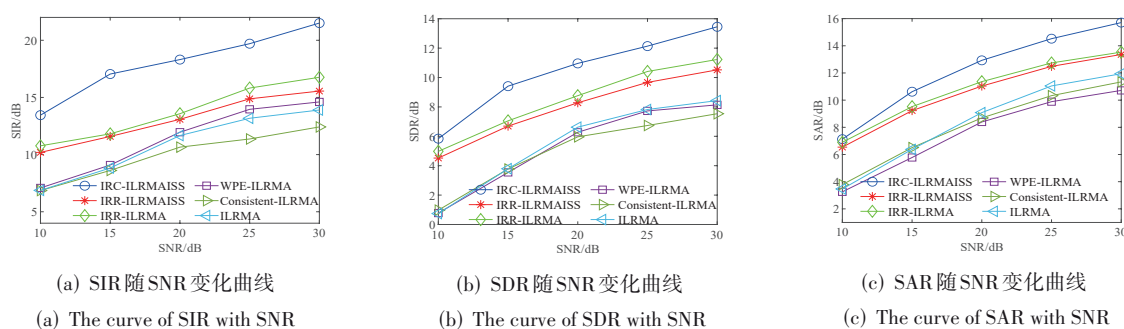


图7 基于音乐信号的盲源分离性能与SNR的关系

Figure 7 The relationship between blind source separation performance and SNR based on music signals

4.3 语音信号实验

选择国际上公开的语音信号数据集 SiSEC'08 中的一路女生英文语音信号和一路男生英文语音信号, 其他参数设置与 4.1 节保持一致。仿真结果如图 8、图 9 和图 10 所示。

由图 8(a)~(c) 可知, 两路语音信号在高混响情况下依然被有效分离, 此时得到的 SIR、SDR 和 SAR 分别为 16.884 2 dB、12.399 3 dB 和 15.759 1 dB。图 9(a)~(c) 分别为针对语音信号分离的 SIR、SDR 和 SAR 随混响

时间变化曲线。图 10(a)~(c) 分别为针对语音信号分离的 SIR、SDR 和 SAR 随 SNR 变化曲线。由仿真结果图 9 和图 10 可知, IRR-ILRMA 和 WPE-ILRMA 的性能均优于没有采取混响抑制手段的 ILRMA 和 Consistent-ILRMA。与动物声音信号和音乐信号不同, 男女语音信号成分存在较大差异, 因此男女语音信号的分离问题相比于 4.1 节和 4.2 节更简单。在 WPE-ILRMA 对应的参考文献[18]中, 也是以语音信号作为源信号进行仿真实验, 并未进行关于动物信号和音乐信号的分

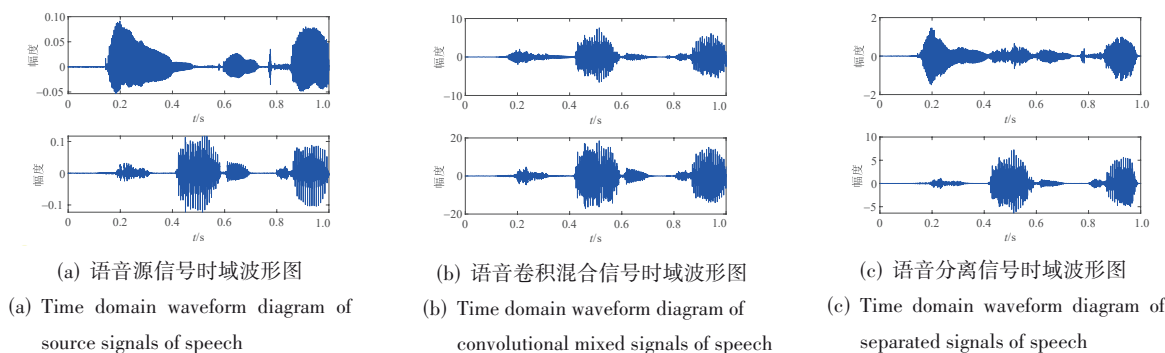


图8 高混响下基于语音信号的 IRC-ILRMAISS 盲源分离效果图

Figure 8 Blind source separation effect diagram of IRC-ILRMAISS based on speech signals under high reverberation

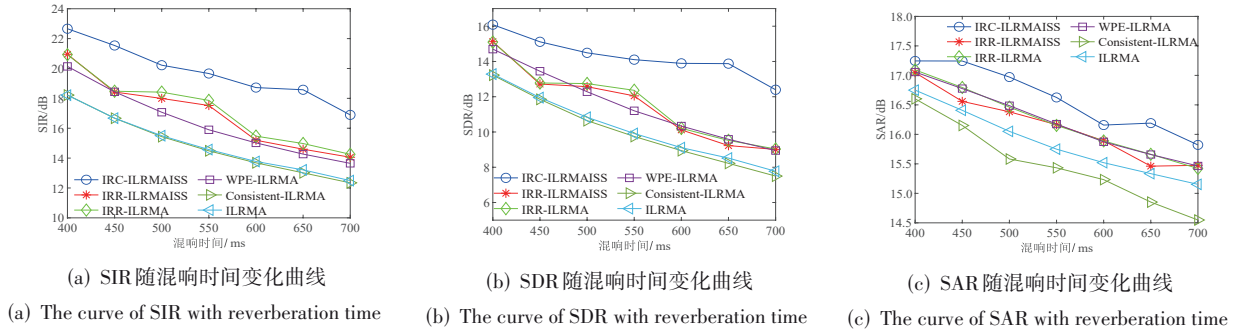


图9 基于语音信号的盲源分离性能与混响时间的关系

Figure 9 The relationship between blind source separation performance and reverberation time based on speech signals

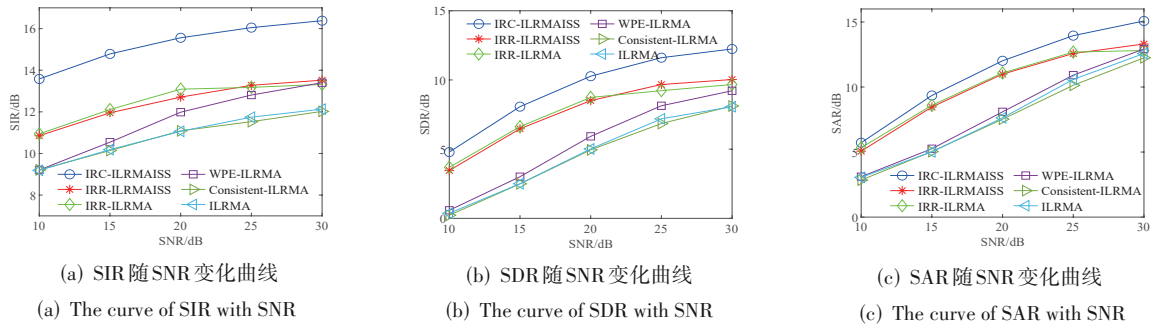


图10 基于语音信号的盲源分离性能与 SNR 的关系

Figure 10 The relationship between blind source separation performance and SNR based on speech signals

离实验。因此区别于4.1节和4.2节的仿真结果,在无噪声影响的情况下,即图9(a)~(c)中,WPE-ILRMA和IRR-ILRMA的性能接近,但是当有噪声影响时,即图10(a)~(c)中,WPE-ILRMA的性能则下降,这一现象与4.1节WPE-ILRMA抗噪能力不足的结论相符合。然而,在高混响环境下,不论是针对成分差异较小的动物声音信号分离问题、成分复杂的音乐信号分离问题,还是成分差异较大的语音信号分离问题,所提出的IRC-ILRMAISS相比于其他方法均实现了更优越的BSS性能,这表明所提出方法拥有较强的鲁棒性,应用范围广泛。

4.4 窗函数敏感度分析

为了测试窗函数设置对混响抑制性能的影响,比较在所提方法、文献[20]和文献[21]的窗函数设置方式下的IRC-ILRMAISS的性能,选择4.2节中的两路音乐信号作为声源信号,考虑无噪声影响且混响时间为700 ms的情况,在每种窗函数设置方式下进行50次仿真实验,取平均值作为实验结果,仿真结果如表1所示。

由表1可知,将所提方法的窗函数设置方式更换为已有IRR方法中的方式并不会严重影响IRC的混响抑制性能,并且在文献[20]和文献[21]中均未分

析窗函数设置方式对方法性能的影响。

4.5 仿真时间对比

两路声源信号为4.2节中的音乐信号,混响时间为700 ms,不考虑噪声的影响,其他参数设置与4.1节保持一致,仿真时间结果取50次仿真实验的平均值,如表2所示。

由表2可知,ILRMAISS与传统的ILRMA相比平均仿真时间略提升,Consistent-ILRMA的平均仿真时间最长,ILRMAISS是以牺牲实时性规避了传统ILRMA中大量的矩阵求逆计算,ISS的引入从仿真结果来看并无明显意义,但已有文献表明矩阵求逆计算会消耗硬件资源并带来算法的不稳定性,因此规避矩阵求逆计算有助于算法的硬件实现^[16,19,24]。另外,可以看到当引入去混响方法后,方法的平均仿真时间均大幅提升,这是因为IRC、IRR和WPE均需要迭代计算才能实现高混响的抑制,将IRC和ILRMA结合的平均仿真时间也进一步缩短。综上所述,ISS的引入虽然可以规避矩阵求逆计算,但会造成方法实时性的损失,并且在高混响环境下,所提出的IRC-ILRMAISS相比于IRR-ILRMA的平均仿真时间略短,相比于WPE-ILRMA的平均仿真时间则大幅减少。

表 1 不同窗函数设置方式下 IRC-ILRMAISS 性能对比

窗函数设置方式来源	参数设置具体值	性能评价指标	指标计算值/dB
所提方法	$\alpha_1=\alpha_2=0.004$ $\bar{w}_0(l)=\frac{\beta-1}{L_3-1}l+1$ $\beta=2$	SIR	16.456 4
		SDR	12.446 4
		SAR	16.064 3
文献[20]	$\alpha_1=t_0$ t_0 代表声源有效时间 $\alpha_2=0.004$ $\frac{3\left[\log_{10}\left(\frac{1}{\gamma}\left(L_s-L_1\right)\right)\right]}{\log_{10}\left(L_s\left(L_s-L_1\right)\right)}+\frac{1}{2}$ $\bar{w}_0(l)=10$ $L_0=(0.2+t_0)f_s$	SIR	17.379 7
		SDR	13.181 4
		SAR	16.372 8
文献[21]	$\alpha_1=t_0$ $\alpha_2=0.05$ $\bar{w}_0(l)=\frac{\beta-1}{L_3-1}l+1$ $\beta=2$	SIR	17.305 5
		SDR	13.158 0
		SAR	16.457 3

表 2 仿真时间对比

是否引入去混响方法	算法	平均仿真时间/s
否	ILRMAISS	2.538 8
	ILRMA	2.404 1
	Consistent-ILRMA	3.476 2
是	IRC-ILRMAISS	20.517 5
	IRC-ILRMA	20.466 1
	IRR-ILRMA	20.995 0
	WPE-ILRMA	33.678 0

4.6 多声源信号实验

以上实验均为两路声源信号的卷积盲分离,为了进一步测试多声源信号卷积盲分离场景下的方法性能,从语音信号数据集 SiSEC’08 中选取四路声源信号进行实验,所提方法的 BSS 效果图如图 11 所示。考虑

无噪声影响且混响时间为 700 ms 的情况,指标计算值结果取 50 次仿真实验的平均值,仿真结果如表 3 所示。四个声源位置分别为[0.5 0 0.36]、[0.5 0.3 0.36]、[0.5 0.6 0.36]和[0.5 0.9 0.36],四个扩音器位置分别为[0.6 0.05 0.26]、[0.6 0.25 0.26]、[0.6 0.45 0.26]和[0.6 0.65 0.26],四个麦克风位置分别为[0.7 0.1 0.16]、[0.7 0.2 0.16]、[0.7 0.3 0.16]和[0.7 0.4 0.16],其他参数设置与 4.1 节保持一致。

由图 11(a)~(c)可知,四路语音信号在高混响情况下依然被有效分离,此时得到的 SIR、SDR 和 SAR 分别为 15.144 8 dB、11.894 6 dB 和 15.933 5 dB。由表 3 可知,当信源数增多时,所有方法性能均略有下降,没有抑制混响的 Consistent-ILRMA 和 ILRMA 性能依然不如其他方法,所提出的 IRC-ILRMAISS 相比于其他方法性能依然更优越。

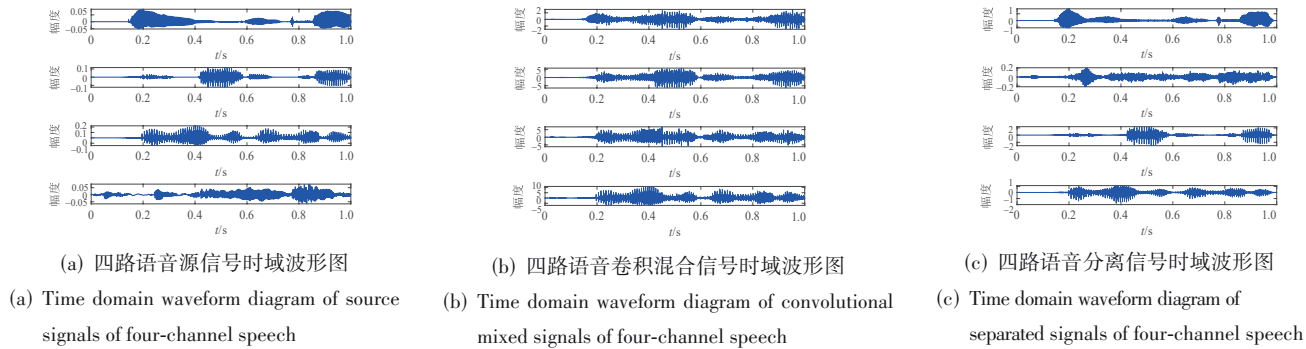


图 11 高混响下基于四路语音信号的 IRC-ILRMAISS 盲源分离效果图

Figure 11 Blind source separation effect diagram of IRC-ILRMAISS based on four-channel speech signals under high reverberation

表3 四路声源信号卷积盲分离场景下的方法性能对比
Table 3 Performance comparison of four-channel convolutional blind source separation for audio signals

算法	性能评价指标	指标计算值/dB
IRC-ILRMAISS	SIR	14.522 4
	SDR	11.160 7
	SAR	15.175 7
IRR-ILRMAISS	SIR	11.147 9
	SDR	8.284 6
	SAR	13.164 4
IRR-ILRMA	SIR	12.651 2
	SDR	9.503 7
	SAR	14.249 0
WPE-ILRMA	SIR	11.509 5
	SDR	8.442 4
	SAR	12.264 5
Consistent-ILRMA	SIR	6.078 1
	SDR	3.948 8
	SAR	10.951 5
ILRMA	SIR	7.153 3
	SDR	5.089 8
	SAR	12.289 2

5 结论

针对现有高混响环境下的卷积盲源分离方法性能依赖 p 范数或滤波器阶数设置的难题,提出了一种基于脉冲响应压缩的卷积盲源分离方法。受传统压缩变换思想启发,设计了一种新的脉冲响应压缩模型,并给出了严格的理论推导,证明了所提出的脉冲响应压缩模型可以在无需设置和调节 p 范数或滤波器阶数的情况下实现有效的高混响抑制。另外,引入迭代源控制算法,在方法实时性牺牲较小的情况下规避了矩阵求逆计算。相比于已有卷积盲源分离方法,所提出方法的性能不依赖于 p 范数值或滤波器阶数的设置和调节,分离精度更高,鲁棒性更强,并且更易于工程实现,突破了现有卷积盲源分离方法的应用局限,所提出的脉冲响应压缩模型也可以移植应用于其他与混响抑制相关的工程问题中。由于所提出方法并未显著降低计算复杂度,并且应用于多通道盲源分离场景,因此未来还将继续研究复杂度较低的去混响方法以及单通道盲源分离方法。

参考文献

- [1] Comon P, Jutten C. Handbook of blind source separation: Independent component analysis and applications[M]. Sea Harbor Drive Orlando: Academic Press, Inc., 2010.
- [2] Gao Hongyuan, Zhang Zhiwei, Zhu Qinglin, et al. High-resolution distributed array DOA estimation based on phase offset recovery between subarrays[J]. Physical Communication, 2025, 71: 102708.
- [3] Wang Lu, Zhao Chunhui, Dong Mianxiong, et al. Fetal ECG signal extraction from long-term abdominal recordings based on adaptive QRS removal and joint blind source separation[J]. IEEE Sensors Journal, 2022, 22(21): 20718-20729.
- [4] Ueda T, Nakatani T, Ikeshita R, et al. Blind and spatially-regularized online joint optimization of source separation, dereverberation, and noise reduction[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024, 32: 1157-1172.
- [5] Liao Xiangyu, Chen Qian, Xiang Jiawei. Underdetermined blind source separation with adaptive chirp mode decomposition for compound rolling bearing fault signals[J]. Measurement Science and Technology, 2025, 36(2): 026123.
- [6] Su Qiao, Zhang Xiongwei, Sha Nan, et al. Underdetermined blind direction-of-arrival estimation using a moving platform[J]. IEEE Signal Processing Letters, 2022, 29: 2532-2536.
- [7] Ma Baoze, Li Guojun, Yi Chen. Tensor-based underdetermined blind identification of instantaneous mixtures[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2023, 70(1): 346-350.
- [8] Wang Taihui, Yang Feiran, Yang Jun. 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.
- [9] Guo Jiayi, Liu Sen, Yu Kun, 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: 9001008.
- [10] Nakatani T, Yoshioka T, Kinoshita K, et al. Speech dereverberation based on variance-normalized delayed linear prediction[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(7): 1717-1731.
- [11] Yoshioka T, Nakatani T, Miyoshi M, et al. Blind separation and dereverberation of speech mixtures by joint optimization[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(1): 69-84.
- [12] 康坊, 杨飞然, 杨军. 子带t分布的快速独立向量分析在语音盲源分离中的应用[J]. 应用声学, 2022, 41(2): 173-181. Kang Fang, Yang Feiran, Yang Jun. Fast independent vector analysis based on clique t-distribution source model for blind speech separation[J]. Journal of Applied Acoustics, 2022, 41(2): 173-181. (in Chinese)
- [13] Kitamura D, Ono N, Sawada H, et al. Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization[J]. IEEE/ACM

- Transactions on Audio, Speech, and Language Processing, 2016, 24(9): 1626-1641.
- [14] Kitamura D, Yatabe K. Consistent independent low-rank matrix analysis for determined blind source separation[J]. EURASIP Journal on Advances in Signal Processing, 2020, 2020(1): 46.
- [15] 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.
- [16] Nakashima T, Scheibler R, Togami M, et al. Joint dereverberation and separation with iterative source steering[C]// ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2021: 216-220.
- [17] Févotte C, Bertin N, Durrieu J L. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis[J]. Neural Computation, 2009, 21(3): 793-830.
- [18] 王泰辉, 杨飞然, 杨军. 低复杂度的盲源分离和去混响联合优化方法[J]. 声学学报, 2024, 49(1): 163-170.
Wang Taihui, Yang Feiran, Yang Jun. A low-complexity joint optimization of blind source separation and dereverberation[J]. Acta Acustica, 2024, 49(1): 163-170. (in Chinese)
- [19] Jin Long, Li Shuai, Hu Bin. RNN models for dynamic matrix inversion: A control-theoretical perspective[J]. IEEE Transactions on Industrial Informatics, 2018, 14(1): 189-199.
- [20] 解元, 张旭, 邹涛, 等. 结合脉冲响应重塑和期望最大化的盲信号分离[J]. 电子学报, 2023, 51(11): 3343-3353.
Xie Yuan, Zhang Xu, Zou Tao, et al. Blind signal separation combining impulse response remodeling and expectation maximization[J]. Acta Electronica Sinica, 2023, 51(11): 3343-3353. (in Chinese)
- [21] 解元, 邹涛, 孙为军, 等. 面向高混响环境的欠定卷积盲源分离算法[J]. 通信学报, 2023, 44(2): 82-93.
Xie Yuan, Zou Tao, Sun Weijun, et al. Algorithm of underdetermined convolutive blind source separation for high reverberation environment[J]. Journal on Communications, 2023, 44(2): 82-93. (in Chinese)
- [22] Zhang Zhiwei, Gao Hongyuan, Zhu Qinglin, et al. Blind source separation method based on blind compression transformation under impulsive noise[J]. Digital Signal Processing, 2025, 161: 105095.
- [23] 高洪元, 张志伟, 刘亚鹏, 等. 一种基于脉冲响应压缩的卷积盲源分离方法、系统及存储介质: CN202511664377.1[P]. 2025-12-30.
Gao Hongyuan, Zhang Zhiwei, Liu Yapeng, et al. Convolutional blind source separation method and system based on impulse response compression and storage medium: CN202511664377.1[P]. 2025-12-30.
- [24] Scheibler R, Ono N. Fast and stable blind source separation with rank-1 updates[C]//ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2020: 236-240.
- [25] Vincent E, Gribonval R, Févotte C. Performance measurement in blind audio source separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1462-1469.
- [26] Yang Junjie, Guo Yi, Yang Zuyuan, et al. Under-determined convolutive blind source separation combining density-based clustering and sparse reconstruction in time-frequency domain[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2019, 66(8): 3015-3027.

作者简介



张志伟 男, 1996年出生于安徽省淮北市。现为哈尔滨工程大学博士研究生。主要研究方向为盲源分离、智能计算理论及应用。
E-mail: zhangzhiwei12138@hrbeu.edu.cn



王佳仪 女, 2002年出生于黑龙江省鹤岗市。现为哈尔滨工程大学博士研究生。主要研究方向为阵列信号处理、智能计算理论及应用。
E-mail: wangjiayi@hrbeu.edu.cn



高洪元 男, 1977年出生于黑龙江省木兰县。现为哈尔滨工程大学副教授、博士生导师。主要研究方向为无线通信系统、阵列信号处理、智能计算理论及应用。中国电子学会会员编号: E190014628M。
E-mail: gaohongyuan@hrbeu.edu.cn



刘亚鹏 男, 1999年出生于河北省邯郸市。现为哈尔滨工程大学博士研究生。主要研究方向为阵列信号处理、智能计算理论及应用。
E-mail: liuyapeng@hrbeu.edu.cn