

基于张量分解的卷积盲源分离方法

马宝泽, 张天骐, 安泽亮, 邓盼

(重庆邮电大学通信与信息工程学院信号与信息处理重庆市重点实验室, 重庆 400065)

摘 要: 基于张量分解框架提出了一种卷积盲源分离方法, 同时解决了混合滤波器矩阵估计和频点排序的问题。首先, 根据观测信号的估计自相关矩阵构造出所有频点处的张量模型; 然后, 利用张量分解技术计算出每个频点上对应的因子矩阵作为该频点的估计混合滤波器矩阵; 最后, 采用以功率比作为测度的全局优化排序策略消除了全频段的排序模糊性。实验表明, 所提方法在不同仿真条件下处理卷积混合的实测语音时表现出了比现有算法更优异的分离性能。

关键词: 卷积盲源分离; 张量分解; 自相关矩阵; 排序模糊

中图分类号: TN911.6

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021140

Convolutional blind source separation method based on tensor decomposition

MA Baoze, ZHANG Tianqi, AN Zeliang, DENG Pan

Chongqing Key Laboratory of Signal and Information Processing, School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: A convolutional blind source separation algorithm was proposed based on tensor decomposition framework, to address the estimation of mixed filter matrix and the permutation alignment of frequency bin simultaneously. Firstly, the tensor models at all frequency bins were constructed according to the estimated autocorrelation matrix of the observed signals. Secondly, the factor matrix corresponding to each frequency bin was calculated by tensor decomposition technique as the estimated mixed filter matrix for that bin. Finally, a global optimal permutation strategy with power ratio as the permutation alignment measure was adopted to eliminate the permutation ambiguity in all the frequency bins. Experimental results demonstrate that the proposed method achieves better separation performance than other existing algorithms when dealing with convolutional mixed speech under different simulation conditions.

Keywords: convolutional blind source separation, tensor decomposition, autocorrelation matrix, permutation ambiguity

1 引言

盲源分离 (BSS, blind source separation) 是一种在信源和混合信道系统均未知的状况下从观测信号中提取源信号的信号处理技术^[1], 其中考虑多径和时延的卷

积盲源分离 (CBSS, convolutional BSS) 是比瞬时 BSS 更复杂的课题^[2]。在 CBSS 中, 传感器接收到的观测信号是由多维源信号经过衰减时延后组成的, 例如, 在混响环境下采集的语音信号^[3], 会导致现有的瞬时 BSS 方法的失效。此外, 在信号传输过程中, 卷积混合比

收稿日期: 2021-04-22; 修回日期: 2021-06-30

基金项目: 国家自然科学基金资助项目 (No.61671095, No.61371164); 信号与信息处理重庆市市级重点实验室建设基金资助项目 (No.CSTC2009CA2003); 重庆市教育委员会科研基金资助项目 (No.KJ130524, No.KJ1600427, No.KJ1600429)

Foundation Items: The National Natural Science Foundation of China (No.61671095, No.61371164), The Project of Key Laboratory of Signal and Information Processing of Chongqing (No.CSTC2009CA2003), The Research Project of Chongqing Educational Commission (No.KJ130524, No.KJ1600427, No.KJ1600429)

瞬时混合更符合实际情况，因此从卷积混合的多维观测信号中分离出源信号是很有挑战性的工作。

由于解决 CBSS 的时域方法受卷积长度影响较大，且存在计算复杂度大、收敛速度慢的缺点，因而频域方法得到了广泛的研究^[2-4]。频域方法首先通过时频变换将时域卷积混合模型转换为每个频点上的频域瞬时混合模型，然后利用复独立分量分析 (CICA, complex independent component analysis) 算法处理各个频点上的序列，最后结合所有频点重构时域信源。即使在每个频点上都能得到理想的瞬时分离结果，但排序和幅度的模糊性也会影响源信号的恢复^[3-5]。文献[3]阐述了联合块对角化和块 Tucker 分解之间的关系，解决了频谱分离和频点排序模糊性问题。文献[4]介绍了一种低复杂度的分步排序方法，权衡了分离性能和计算复杂度的矛盾，但没有提出新的混合滤波器矩阵估计方案。文献[5]将二元经验模态分解与熵界最小化复独立分量分析相结合实现了语音信号的卷积盲分离，但缺乏解决排序模糊性方面的工作量。此外，独立向量分析 (IVA, independent vector analysis)^[6]提出了引入矢量源模型的方式来解决卷积盲信号分离问题。此后，独立低秩矩阵分析 (ILRMA, independent low-rank matrix analysis)^[7]整合了 IVA 和非负矩阵分解，用低秩时频结构的源模型代替了 IVA 中的矢量源模型。虽然 IVA 和 ILRMA 可以实现卷积混合信号的分离，但它们对信源模型的假设严重限制了算法的适用范围。因此，本文将针对 CBSS 频域方法展开研究，同时解决了混合滤波器矩阵估计和频点排序 2 个问题，实现了全面提高 CBSS 算法性能的目标。

由于多维信号的频谱可以容易地表示成张量形式，同时经张量分解后得到的因子矩阵也可以用于估计混合滤波器矩阵。因此，本文在张量分解的框架下研究了一种 CBSS 频域方法。首先，将观测信号序列分为若干非重叠子块，并对每个子块进行时频变换得到张量形式的估计自相关矩阵。其次，根据张量分解技术估计出与每个频点相对应的混合滤波器矩阵，即解决每个频点上的频谱分离问题。再次，采用全局优化的频点排序策略，将功率比作为排序测度保障了属于相同信源的频谱被归为同一类。最后，排序后的全部频点经逆变换得到时域分离信号。仿真实验中以实测语音作为信源，在模拟的立体空间中产生房间脉冲响应实现卷积混合。此外，本文方法相较现有方法的有效性和优

越性在多种实验条件下得到了验证。

2 问题描述

CBSS 混合系统可以用长度为 L 的有限脉冲响应 (FIR, finite impulse response) 滤波器表示，那么 n 个源信号 $\mathbf{s}(t)=[s_1(t), s_2(t), \dots, s_n(t)]^T$ 经过信道的线性卷积混合后可以得到 m 个观测信号，即 $\mathbf{x}(t)=[x_1(t), x_2(t), \dots, x_m(t)]^T$ 。不考虑噪声的卷积混合模型可以表示为

$$\mathbf{x}(t) = \mathbf{H} * \mathbf{s}(t) = \sum_{l=0}^{L-1} \mathbf{H}(l) \mathbf{s}(t-l) \quad (1)$$

其中， $*$ 表示线性卷积运算， $\mathbf{H}(l) \in \mathbb{R}^{m \times n \times L}$ 是第 l 个脉冲响应混合系统， $l=0, 1, 2, \dots, L-1$ ， L 是混合信道系统的阶数。此外， $\mathbf{H}(l)$ 可以表示成矩阵的形式，即

$$\mathbf{H}(l) = \begin{bmatrix} h_{11}(l) & h_{12}(l) & \cdots & h_{1n}(l) \\ h_{21}(l) & h_{22}(l) & \cdots & h_{2n}(l) \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1}(l) & h_{m2}(l) & \cdots & h_{mn}(l) \end{bmatrix} \quad (2)$$

其中， $h_{ji}(l)$ 表示从第 i 个信源到第 j 个传感器的房间脉冲响应 (RIR, room impulse response)，这个过程可以用一个 FIR 滤波器来模拟。CBSS 就是在混合信道系统 \mathbf{H} 未知的情况下，通过寻找一个 $n \times m$ 的近似可逆的信道系统 \mathbf{w} 来实现信源的估计，那么分离信号可以表示为

$$\mathbf{y}(t) = \mathbf{w} * \mathbf{x}(t) = \sum_{k=0}^{K-1} \mathbf{w}(k) \mathbf{x}(t-k) \quad (3)$$

其中， K 是解混信道系统的总阶数。分离信号 $\mathbf{y}(t)=[y_1(t), y_2(t), \dots, y_n(t)]^T$ 是源信号 $\mathbf{s}(t)$ 的估计，但是 $\mathbf{y}(t)$ 和 $\mathbf{s}(t)$ 之间存在由 BSS 模糊性造成的幅度和排序差异。

CBSS 频域方法需要利用离散傅里叶变换 (DFT, discrete Fourier transform) 将时域信号变换到频域^[2-5]，式(1)可以表示为

$$\mathbf{X}(f, p) = \mathbf{H}(f) \mathbf{S}(f, p) \quad (4)$$

其中， $\mathbf{X}(f, p)=[X_1(f, p), X_2(f, p), \dots, X_m(f, p)]^T$ 和 $\mathbf{S}(f, p)=[S_1(f, p), S_2(f, p), \dots, S_n(f, p)]^T$ 分别表示观测信号和源信号的时频谱， $\mathbf{H}(f)$ 的第 i 列表示第 i 个信源在频点 f 上的空间特征信息。此外， $f=1, 2, \dots, F$ 表示频点的索引， F 是频点总数， $p=1, 2, \dots, P$ 表示时间帧的索引， P 是时间帧总数。

由式(4)可知, CBSS 问题在时频域被转化为了 F 个瞬时混合 BSS 问题, 即可以利用 CICA 方法处理每个频点上的频谱序列, 然后对频点进行排序, 最后通过快速傅里叶逆变换 (IFFT, inverse fast Fourier transform) 恢复时域分离信号。

3 估计信道构造张量模型

假设观测信号的长度为 N , 那么将观测序列数据分为 P 个非重叠子块, 这样每个子块就包含 $N_b = \lfloor N/P \rfloor$ 个点, 这些子块可以用 $p=1, 2, \dots, P$ 来索引, 第 p 个子块包含的数据序列从 t_{p-1} 到 t_p [3]。因此, 观测信号在第 f 个频点和第 p 个子块处的自相关矩阵可以用 $\mathbf{R}_x(f, p) \in \mathbb{C}^{m \times m}$ 表示, 也就是说, $\mathbf{R}_x(f, p) = E[\mathbf{X}(f, p)\mathbf{X}^H(f, p)]$, 即

$$\mathbf{R}_x(f, p) = \mathbf{H}(f)\mathbf{R}_s(f, p)\mathbf{H}^H(f) \quad (5)$$

其中, $\mathbf{R}_s(f, p) = E[\mathbf{S}(f, p)\mathbf{S}^H(f, p)]$ 是源信号在第 f 个频点和第 p 个子块处的自相关矩阵, $(\cdot)^H$ 表示 Hermitian 转置运算。

不同于将时间帧依次处理, 本文在给定的频点处同时研究 P 个子块, 即

$$\begin{cases} \mathbf{R}_x(f, 1) = \mathbf{H}(f)\mathbf{R}_s(f, 1)\mathbf{H}^H(f) \\ \mathbf{R}_x(f, 2) = \mathbf{H}(f)\mathbf{R}_s(f, 2)\mathbf{H}^H(f) \\ \vdots \\ \mathbf{R}_x(f, P) = \mathbf{H}(f)\mathbf{R}_s(f, P)\mathbf{H}^H(f) \end{cases} \quad (6)$$

对每个频点上的 $\mathbf{H}(f)$ 进行估计就相当于处理一个联合相似对角化 (JAD, joint approximate diagonalization) 问题, 那么式(6)中就存在 P 个 JAD 问题, 将它们看作一个整体, 借助张量的思想解决。这样就建立起 JAD 和张量分解之间的关系, 即将每个频点上的传统 JAD 问题转变成张量分解问题。然而, 利用张量分解得到的因子矩阵估计混合滤波器矩阵的关键在于自相关矩阵的计算。实际上, 人们很难获得准确的观测信号自相关矩阵 $\mathbf{R}_x(f, p)$, 只能通过时域观测信号 $\mathbf{x}(t), t=1, 2, \dots, N$ 来估计。在含有 N_b 个采样点的每个子块中, 选择汉宁窗对若干连续重叠帧计算 N_f 点 DFT, 其中, 帧长和窗长都为 N_f , 2 个连续帧之间重叠为窗长的 3/4, 即 $\text{ovlp} = 75\%$ [3,4]。因此, 每个子块中的重叠帧数可以表示为

$$N_t = \left\lfloor \frac{N_b - N_f}{N_f(1 - \text{ovlp})} \right\rfloor + 1 \quad (7)$$

其中, $\lfloor \cdot \rfloor$ 表示向零方向取整运算。由于实信号的频谱是对称的, 因此有用的频点数为 $F = N_f/2 + 1$ 。那么, 第 f 个频点和第 p 个子块处对应的估计自相关矩阵可以表示为

$$\hat{\mathbf{R}}_x(f, p) = \frac{1}{N_t} \sum_{c=1}^{N_t} \left[\mathbf{X}(f, k_p^c + 1 : k_p^c + N_f) \cdot \mathbf{X}^H(f, k_p^c + 1 : k_p^c + N_f) \right] \quad (8)$$

其中, $k_p^c = (p-1)N_b + (c-1)N_f(1 - \text{ovlp})$ 是第 p 个子块中第 c 个重叠帧的索引, $f=1, 2, \dots, F$ 。

根据上述分析, 观测信号的第 f 个频点上可以计算出 P 个估计自相关矩阵, 可以表示为集合 $\{\hat{\mathbf{R}}_x(f, 1), \hat{\mathbf{R}}_x(f, 2), \dots, \hat{\mathbf{R}}_x(f, P)\}$, 然后, 将该集合中的元素一个接一个地保存就可以得到一个三维张量 $\mathfrak{R}_x(f) \in \mathbb{C}^{m \times m \times P}$ 。由于式(6)相当于一个张量模型, 那么估计自相关矩阵构成的张量 $\mathfrak{R}_x(f)$ 中的每个元素可以表示为 $r_x^{j_1, j_2, P}(f)$, $j_1=1, 2, \dots, m$, $j_2=1, 2, \dots, m$, $p=1, 2, \dots, P$ 。 $h_{ji}(f)$ 是 $\mathbf{H}(f)$ 的元素, 表示在第 f 个频点处从第 i 个信源到第 j 个传感器的信道信息。此外, 定义一个 $P \times n$ 的矩阵 $\mathbf{G}(f)$ 作为 $\mathbf{R}_s(f, p)$ 的对角矩阵, 那么 $\mathbf{R}_s(f, p)$ 的第 i 个对角元素可以用 $\mathbf{G}(f)$ 的第 p 行、第 i 列元素 $g_{pi}(f)$ 表示。因此, $r_x^{j_1, j_2, P}(f)$ 可以表示为张量 $\mathfrak{R}_x(f)$ 的共轭对称的分解形式, 即

$$r_x^{j_1, j_2, P}(f) = \sum_{i=1}^n h_{ji}(f) g_{pi}(f) h_{ji}^*(f) \quad (9)$$

其中, $(\cdot)^*$ 表示复共轭运算。由式(9)可知, 张量 $\mathfrak{R}_x(f)$ 的秩等于信源数 n 。由于张量的所有维度必须大于或等于秩 [8], 因此必须满足条件 $m \geq n$ 和 $P \geq n$ 。此外, 只有每个子块中语音信号的音长 $T_p = N_b/f_s$ 满足 $T_p \geq 40$ ms 时才能保证语音信号的非平稳性 [3], 其中 f_s 是采样频率。为了达到音长的最低标准, 本文在 $n \leq m \leq P$ 的条件下实现卷积信号盲分离。

4 张量分解估计混合滤波器矩阵

由于时域卷积混合 BSS 在时频变换后就转变成了时频域瞬时混合 BSS 问题, 那么第 f 个频点对应张量 $\mathfrak{R}_x(f)$ 的分解就相当于实现了复数形式的瞬时混合盲信号分离。因此, 通过张量分解的方式解决每个频点的频谱分离问题是值得研究的。由上

述分析可知，本文处理的是三维张量，在张量分解前需要调整各维度顺序。 $\mathfrak{R}_x(f)$ 的第三维 P 是张量维度中最大的，这样就保证了三维张量的维度是升序状态。为了消除非重叠子块 P 选取过大对张量分解的影响，将张量的第三维由 P 降维到 $n+1$ ，即 $\mathfrak{R}_x(f) \in \mathbb{C}^{m \times m \times P}$ 转变为 $\mathfrak{R}_x(f) \in \mathbb{C}^{m \times m \times (n+1)}$ 。然后，利用扩展 QZ 分解对降维后的张量 $\mathfrak{R}_x(f)$ 进行处理，实现舒尔 (Schur) 上三角化^[9]。

选取 $\mathfrak{R}_x(f)$ 沿着第三维的前 2 个切片 (Slice) 矩阵 $\mathfrak{R}_x^{(1)}(f) \in \mathbb{C}^{m \times m}$ 和 $\mathfrak{R}_x^{(2)}(f) \in \mathbb{C}^{m \times m}$ 进行复数域广义特征值的 QZ 分解，作为张量分解的初始化。

$$\begin{cases} \mathbf{Q}(f)\mathfrak{R}_x^{(1)}(f)\mathbf{Z}(f) = \mathbf{U}_1(f) \\ \mathbf{Q}(f)\mathfrak{R}_x^{(2)}(f)\mathbf{Z}(f) = \mathbf{U}_2(f) \end{cases} \quad (10)$$

其中， $\mathbf{Q}(f) \in \mathbb{C}^{m \times m}$ 和 $\mathbf{Z}(f) \in \mathbb{C}^{m \times m}$ 、 $\mathbf{U}_1(f) \in \mathbb{C}^{m \times m}$ 和 $\mathbf{U}_2(f) \in \mathbb{C}^{m \times m}$ 分别是 $\mathfrak{R}_x^{(1)}(f)$ 、 $\mathfrak{R}_x^{(2)}(f)$ 分解后对应的上三角矩阵。此时的 $\mathbf{Q}(f)$ 和 $\mathbf{Z}(f)$ 是通过 2 个张量切片初始化的，接下来，需要将它们迭代更新。首先，计算张量 $\mathfrak{R}_x(f)$ 和矩阵 $\mathbf{Q}(f)$ 的 1-模态积 (mode-1 product)。然后，将结果与矩阵 $\mathbf{Z}(f)$ 的转置做 2-模态积 (mode-2 product) 运算，得到解析张量 $\mathfrak{S}(f) \in \mathbb{C}^{n \times n \times (n+1)}$ 。

$$\mathfrak{S}(f) = [\mathfrak{R}_x(f) \times_1 \mathbf{Q}(f)] \times_2 \mathbf{Z}^T(f) \quad (11)$$

其中， x_1 和 x_2 分别表示 1-模态积和 2-模态积。最后，更新 $\mathbf{Q}(f)$ 和 $\mathbf{Z}(f)$ ，根据式(11)更新 $\mathfrak{S}(f)$ ，直到达到迭代终止条件。

通过更新后的 $\mathbf{Q}(f)$ 和 $\mathbf{Z}(f)$ 估计混合滤波器 $\mathbf{h}(f)$ ，将张量 $\mathfrak{S}(f)$ 按照第三维方向的切片得到的上三角矩阵对角线作为矩阵 $\Psi(f) \in \mathbb{C}^{(n+1) \times n}$ 的列；再计算张量 $\mathfrak{R}_x(f)$ 和矩阵 $\Psi(f)$ 逆的 3-模态积，即

$$\mathfrak{N}(f) = \mathfrak{R}_x(f) \times_3 \Psi(f)^{-1} \quad (12)$$

其中， $\mathfrak{N}(f) \in \mathbb{C}^{n \times m \times n}$ 是估计混合滤波器的解析张量。最后选取 $\mathfrak{N}(f)$ 按照第三维方向的 n 个切片经奇异值分解后的左奇异矩阵 (因子矩阵) 的第一列作为估计混合滤波器矩阵的列，即可以表示为 $\hat{\mathbf{h}}_i(f) = [\hat{h}_{1i}, \hat{h}_{2i}, \dots, \hat{h}_{mi}]^T$, $i=1, 2, \dots, n$ 。根据上述张量分解的步骤，第 f 个频点对应的估计混合滤波器矩阵 $\hat{\mathbf{H}}(f) = [\hat{\mathbf{h}}_1(f), \hat{\mathbf{h}}_2(f), \dots, \hat{\mathbf{h}}_n(f)]$ 就可以计算出

来。以此类推，整个分离过程需要进行 F 次运算。

5 消除幅度和排序的模糊性

理想情况下，分离滤波器矩阵 $\mathbf{W}(f)$ 需要满足 $\mathbf{W}(f)\mathbf{H}(f) = \mathbf{I}$ 的条件，其中， \mathbf{I} 是单位矩阵。那么，在第 f 个频点处就可以根据 $\mathbf{H}(f)$ 求出 $\mathbf{W}(f)$ ，从而得到相应的分离信号的频谱。

$$\mathbf{Y}(f, p) = \mathbf{W}(f)\mathbf{X}(f, p) = \mathbf{A}(f)\mathbf{D}(f)\mathbf{S}(f, p) \quad (13)$$

其中， $\mathbf{Y}(f, p) = [Y_1(f, p), Y_2(f, p), \dots, Y_n(f, p)]^T$ 表示分离信号的频谱， $\mathbf{A}(f)$ 和 $\mathbf{D}(f)$ 分别表示第 f 个频点的对角矩阵和排序矩阵。由式(13)可知，源信号与分离信号在每个频点上的频谱会存在幅度和排序的差异，这就需要在将分离信号频谱逆变换到时域前消除幅度和排序模糊性造成的影响^[5]。

每个频点上的频谱序列需要在保证属于同一源信号的频谱成分被划分为同一类的前提下重新排序，即频点排序是利用属于同一信号的频谱成分会表现出明显的相似性的特点来消除排序模糊性。根据张量分解得到的估计混合滤波器 $\hat{\mathbf{H}}(f)$ ，可以将式(4)改写为

$$\mathbf{X}(f, p) = \hat{\mathbf{H}}(f)\mathbf{Y}(f, p) = \sum_{i=1}^n \hat{\mathbf{h}}_i(f)Y_i(f, p) \quad (14)$$

功率比作为一种有效的频点排序测度，可以表征观测信号中第 i 个分离信号的占比^[4]。那么，第 i 个分离信号在第 p 个子块的功率和所有分离信号的功率之和的比值可以表示为

$$v_i(f, p) = \frac{\|\hat{\mathbf{h}}_i(f)Y_i(f, p)\|^2}{\sum_{k=1}^n \|\hat{\mathbf{h}}_k(f)Y_k(f, p)\|^2} \quad (15)$$

其中， $\|\cdot\|^2$ 表示 2-范数运算。每个信号对应 P 个序列，令 $\mathbf{v}_i^f = [v_i(f, 1), v_i(f, 2), \dots, v_i(f, P)]^T$ 表示第 i 个分离信号在第 f 个频点处的功率比，那么不同频点处的 2 个功率比序列的相关系数可以表示为

$$\rho(\mathbf{v}_i^f, \mathbf{v}_j^g) = \frac{r_{ij}(f, g) - \mu_i(f)\mu_j(g)}{\sigma_i(f)\sigma_j(g)} \quad (16)$$

其中， $\rho(\mathbf{v}_i^f, \mathbf{v}_j^g)$ 是第 i 和 j 个分离信号分别第 f 和 g 个频点处的相关系数， $r_{ij}(f, g) = E[\mathbf{v}_i^f \mathbf{v}_j^g]$ ， $\mu_i(f) = E[\mathbf{v}_i^f]$ ， $\sigma_i(f) = \sqrt{E[(\mathbf{v}_i^f)^2] - \mu_i^2(f)}$ ， $E[\cdot]$

是期望。如果分离信号 i 和 j 来自同一信源，那么 $\rho(\mathbf{v}_i^f, \mathbf{v}_j^g)$ 接近 1；反之， $\rho(\mathbf{v}_i^f, \mathbf{v}_j^g)$ 接近 0。

定义功率比序列在第 k 个信道的质心为

$$\mathbf{c}_k = \frac{1}{F} \sum_{f \in F} \mathbf{v}_i^f \Big|_{i=\Pi_f(k)} \quad (17)$$

其中， $k=1,2,\dots,n$ 是信道索引， Π_f 是第 f 个频点处的排序。式(17)是一种全局优化方式，表明对序列的全频段求质心。然后，利用当前频点的功率比序列与质心序列间的相关系数最大化对排序 Π_f 进行更新。

$$\Pi_f \leftarrow \arg \max_{\Pi} \sum_{k=1}^n \rho(\mathbf{v}_i^f, \mathbf{c}_k) \Big|_{i=\Pi_f(k)} \quad (18)$$

式(17)和式(18)不断迭代更新，直到收敛或达到预设终止条件，此时的排序 Π_f 基本保持不变。第 f 个频点排序完成后，分离滤波器矩阵按照 $\mathbf{W}_p(f) \leftarrow \mathbf{W}_i(f) \Big|_{i=\Pi_f(k)}$ 更新，其中， $\mathbf{W}_p(f)$ 是经 $\mathbf{W}(f)$ 排序后的分离滤波器矩阵。此外，幅度模糊性可以通过最小失真准则 (MDP, minimal distortion principle) 解决^[4-5]，即

$$\mathbf{W}_s(f) = \text{diag}[\mathbf{W}_p^{-1}(f)] \mathbf{W}_p(f) \quad (19)$$

其中， $\text{diag}(\cdot)$ 表示对矩阵主对角线元素的提取运算， $(\cdot)^{-1}$ 表示矩阵逆运算。最终， $\mathbf{W}_s(f)$ 经过逆离散傅里叶变换 (IDFT, inverse discrete Fourier transform) 得到时域的分离矩阵 $\mathbf{w}(t)$ ，通过 $\mathbf{w}(t)$ 对观测信号 $\mathbf{x}(t)$ 进行滤波就可以得到时域分离信号 $\mathbf{y}(t)$ 。

图 1 是本文所提方法流程，其中， $\bar{\mathbf{x}}(t) \in \mathbb{R}^{m \times N_b \times P}$ 表示分块后得到的解析信号， \mathbb{R} 和 \mathbb{C} 分别表示实数和复数。图 1 中包含了该方法中涉及的部分主要变量，并标注了相应的维度和数据类型，这样可以更清晰地表明观测信号在不同阶段的变化情况。此外，图 1 中的张量分解和消除模糊性

两部分的内容详见第 4 节和第 5 节。由于多维信号在频域上很容易转换成张量形式，因此，本文提出的结合张量分解和全局优化排序策略的方法比现有的传统 CBSS 方法更有优势。

6 仿真实验

在多种实验条件下验证本文方法的有效性，分别采用 IVA^[6]、ILRMA^[7]、JAD-based^[10] 和 CICA-EBM^[11] 作为对比方法。为了实验公平，方法中涉及的窗函数都选取 Hanning 窗、帧重叠为窗长的 75%。其中，ILRMA 中涉及的初始基矩阵和激活矩阵均采用随机矩阵。采样频率是 16 kHz 语音信号数据来自文献[12]，包括男声和女生朗读的诗词及句子。在模拟的立体空间中，利用声学的图像方法^[13]生成房间脉冲响应 RIR 对语音信号进行卷积处理，其中，RIR 可以通过设置房间的尺寸、信源和传感器的位置以及混响时间来模拟各种情况。为了模拟真实的卷积环境，实验是在一个长 6 m、宽 4.2 m、高 3 m 的虚拟立体房间里进行^[14]，并固定信源和传感器高度为 1.6 m。此外，所有实验都是在 MATLAB2019a 软件上完成的，计算机配置为 Intel(R) Core(TM) i7-10710U CPU @ 1.10 GHz 1.61 GHz，内存 (RAM) 为 16 GB。

利用输出信干比 (SIR, signal-to-interference ratio) 作为性能指标评价信源分离的精度^[3-5,12]。由于式 (3) 中的第 i 个分离信号可以改写为 $y_i(t) = \sum_{j=1}^m w_{ij} * x_j(t), i=1, \dots, n$ ，那么， $y_i(t)$ 的对应的输出 SIR 可以定义为

$$\text{SIR}_i = 10 \lg \frac{\sum_t y_{ii}^2(t)}{\sum_t \sum_{k \neq i} y_{ik}^2(t)} \quad (20)$$

其中， $y_{ii}(t)$ 表示 $y_i(t)$ 中来自第 i 个信源的部分， $y_{ik}(t)$ 表示串音信号。将由式(20)计算出的 n 个输出 SIR 的平均值作为估计信源的全局性能指标，并以

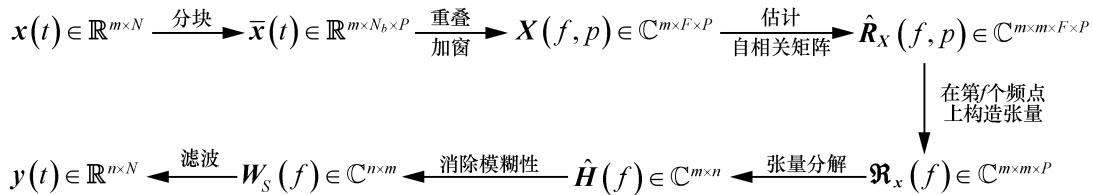


图 1 本文所提方法流程

此作为基准来评估对比算法的分离性能。

6.1 验证 FFT 长度、噪声、混响和信号长度对分离性能的影响

为了验证快速傅里叶变换 (FFT, fast Fourier transform) 长度对 5 种方法分离性能的影响, 选取时长为 5 s 的语音信号作为信源。在 $m = n = 2$ 的情况下, 固定 2 个传感器位置为 (2,1,1.6) 和 (2,2,1.6), 2 个信源位置为 (4,1.5,1.6) 和 (4,2.5,1.6)。此外, 将混响时间 T_{60} 设置为 130 ms^[12]。图 2 是 5 种方法在不同 FFT 长度时的输出 SIR 变化情况。由图 2 可以看出, 当 FFT 长度小于 512 时, 5 种方法的分离性能都较差。随着 FFT 长度的增加, SIR 也随之提高。由于 FFT 长度为 4 096 时各方法的分离性能增长趋势放缓, 甚至 ILRMA 出现了分离性能下降的情况。因此, 综合考虑所有方法对 FFT 长度的需求, 在接下来的实验中都将 FFT 长度设为 2 048。

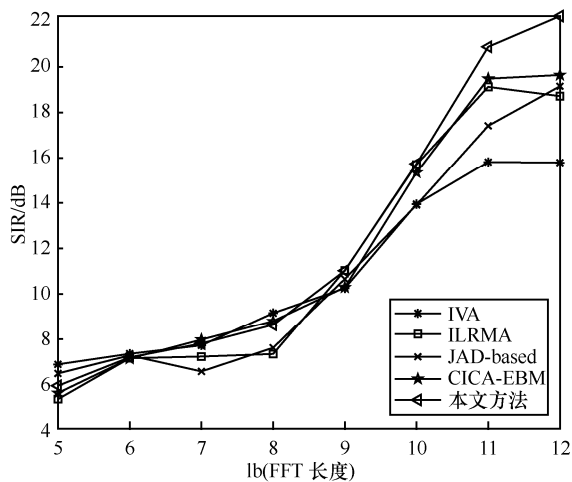


图 2 SIR 随 FFT 长度变化情况

图 3 为 5 种方法在不同信噪比 (SNR, signal-to-noise ratio) 下的稳健性, 随机生成的高斯白噪声使信噪比变化范围由 5 dB 到 60 dB。从图 3 可以看出, 以 CICA-EBM 以及本文方法为代表的 CBSS 频域方法在抗噪性方面要优于 IVA 和 ILRMA。此外, 本文方法在 SNR 大于 15 dB 时表现出了比其他方法更优异的分性能。

混响时间 T_{60} 是影响方法分离性能的重要因素, 混响时间取决于空间尺寸、形状、吸音性能以及信源和传感器的位置。 T_{60} 越大, 表明由多径和时延引起的卷积情况使信号的传播信道变得越复杂。因此, 选取从 0.15 s 到 0.5 s 且间隔为 0.05 s 的时段作为混响区间, 即 $T_{60} \in [0.15, 0.5]$ 。图 4 为混响时间对

5 种方法分离性能的影响。随着 T_{60} 逐渐变大, 所有方法得到的输出 SIR 都随之下降了。在 $m = n = 2$ 的情况下, 本文方法在实验的全时段上的分离性能明显优于其余 4 种对比方法。这说明了本文提出的基于张量分解和功率比全局排序策略的卷积盲源分离方法可有效提升卷积混合语音在强混响情况下的分离性能。

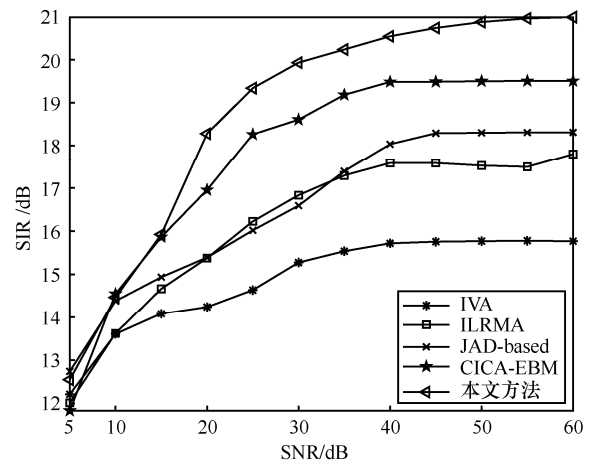


图 3 SIR 随 SNR 变化情况

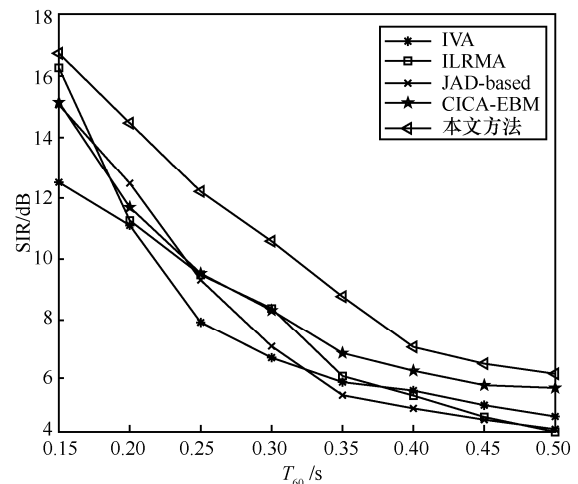
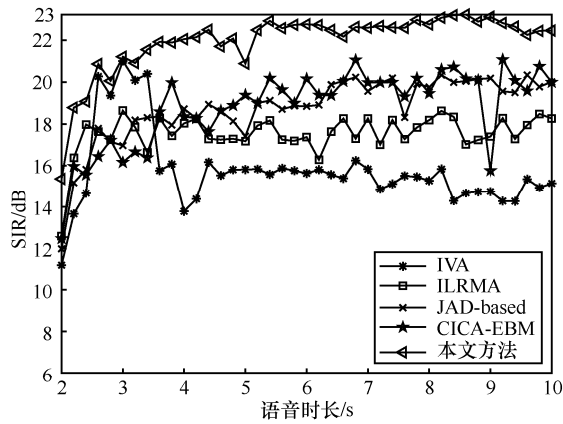
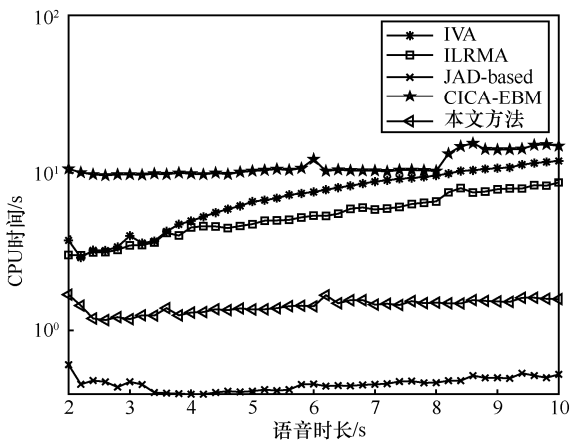


图 4 SIR 随混响时间变化情况

将不同时长的语音作为实验中算法的源信号, 以此验证信号长度对分离性能的影响。选取时长从 2 s 到 10 s 且间隔为 0.2 s 的语音信号进行卷积分离实验, 结果如图 5 所示。图 5(a) 表明各方法的输出 SIR 在不同语音时长下的变化情况, 当信号长度大于 3 s 时, 本文方法比其他 4 种对比方法具有更高的 SIR。此外, 所有方法的计算时间随信号长度变化情况如图 5(b) 所示。可以看出, JAD-based 计算用时最少, 其次是本文方法, 用时最多的是 CICA-EBM。



(a) SIR随信号长度变化情况



(b) CPU时间随信号长度变化情况

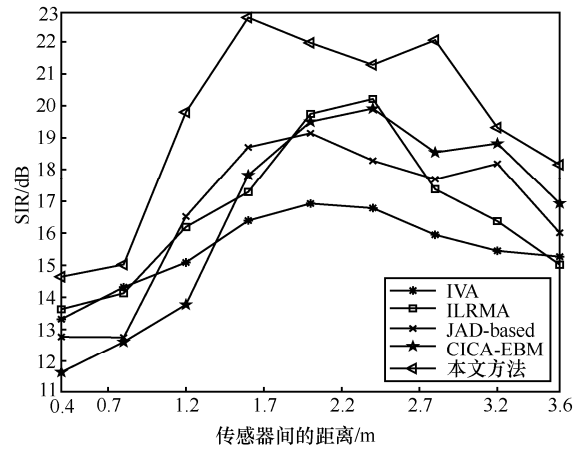
图 5 信号长度对分离性能的影响

6.2 传感器和信源位置对分离性能的影响

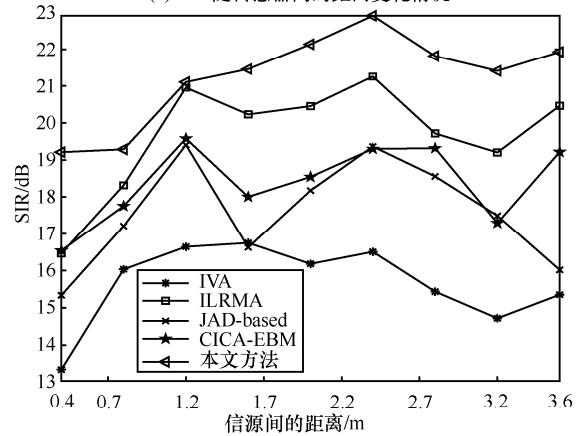
在 $m = n = 2$ 的情况下，验证方法在信源和传感器不同位置下的输出 SIR 情况。根据上述实验结果，将 FFT 长度设置为 2 048、混响时间设置为 130 ms、选取信号时长为 5 s。固定 2 个信源的位置为 (4,1.5,1.6) 和 (4,2.5,1.6)，同时固定一个传感器位置为 (2,0.3,1.6)，另一个传感器以步长为 0.4 m 沿着房间宽度方向移动，这样依次得到了 9 个位置，即 $(2,0.3+0.4k_1,1.6), k_1=1, \dots, 9$ 。通过改变传感器间的距离计算各方法的 SIR，如图 6(a)所示。可以看出，本文方法在 2 个传感器间距离为 1.6 m 时取得最大的 SIR，即移动传感器位置在 (2,1.9,1.6)。

为了验证 2 个信源间距离对算法性能的影响，固定 2 个传感器的位置为 (2,1,1.6) 和 (2,2,1.6)，同时固定一个信源位置为 (4,0.3,1.6)，另一个信源以步长为 0.4 m 沿着房间宽度方向移动，分别得到 9 个位置，即 $(4,0.3+0.4k_2,1.6), k_2=1, \dots, 9$ 。图 6(b)为 2 个信源间的不同距离对

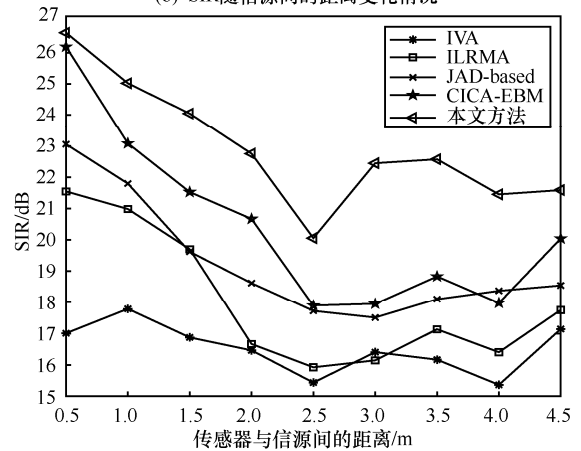
输出 SIR 的影响，且本文方法在 2 个信源间距离为 2.4 m 时得到最大的 SIR，即移动信源的位置为 (4,2.7,1.6)。



(a) SIR随传感器间的距离变化情况



(b) SIR随信源间的距离变化情况



(c) SIR随信源与传感器之间的距离变化情况

图 6 信号长度对分离性能的影响

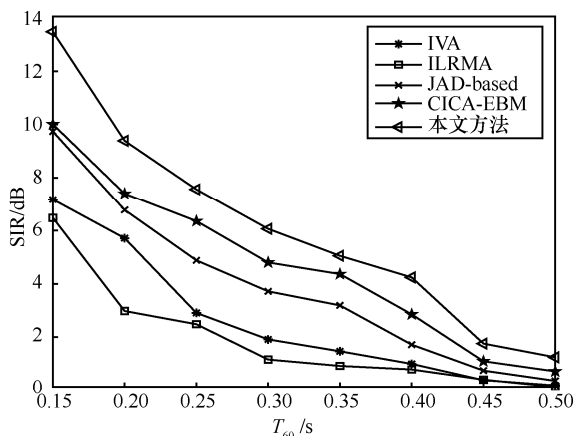
考虑信源与传感器之间的距离对分离性能的影响，固定 2 个传感器的位置为 (1,1,1.6) 和 (1,2,1.6)，同时 2 个信源以步长为 0.5 m 沿着房间长度方向移动，分别得到 2 个信源的 9 个位置，即

一个信源位置可以用 $(1+0.5k_3, 1.5, 1.6)$, $k_3 = 1, \dots, 9$ 表示, 而另一个信源的位置可以用相同的方式表示为 $(1+0.5k_3, 2.5, 1.6)$, $k_3 = 1, \dots, 9$ 。信源和传感器之间的距离由 0.5 m 增加到 4.5 m 的过程中, 5 种方法取得的输出 SIR 如图 6(c) 所示。所有方法基本上都表现出了性能先降后升的趋势, 当距离是 2.5 m 时, 本文方法得到了最小的 SIR。

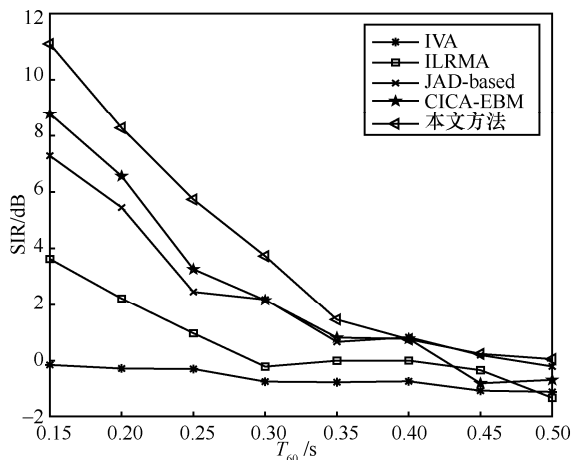
6.3 验证多维信源时混响时间对分离性能的影响

由于源信号和观测信号数目的增加会给 CBSS 带来更大的挑战, 因此, 在 $m = n = 3$ 和 $m = n = 4$ 这 2 种情况下验证混响时间对分离性能的影响。

实验过程中, 除了信源和传感器位置外, 剩余设置都与 6.2 节一致。当 $m = n = 3$ 时, 固定 3 个传感器的位置分别为 $(2, 1, 1.6)$ 、 $(2, 2, 1.6)$ 和 $(2, 3, 1.6)$, 同时固定 3 个信源位置分别为 $(4, 1.5, 1.6)$ 、 $(4, 2.5, 1.6)$ 和 $(4, 3.5, 1.6)$ 。图 7(a) 是 $m = n = 3$ 时输出 SIR 随混响变化的情况, 本文方法分离性能最优, 其次是 CICA-EBM。



(a) $m=n=3$ 时 SIR 随 T_{60} 变化情况



(b) $m=n=4$ 时 SIR 随 T_{60} 变化情况

图 7 混响时间对分离性能的影响

当 $m = n = 4$ 时, 固定 4 个传感器的位置分别为 $(2, 1, 1.6)$ 、 $(2, 2, 1.6)$ 、 $(2, 3, 1.6)$ 和 $(2, 4, 1.6)$, 同时固定 4 个信源位置分别为 $(4, 0.5, 1.6)$ 、 $(4, 1.5, 1.6)$ 、 $(4, 2.5, 1.6)$ 和 $(4, 3.5, 1.6)$ 。实验结果如图 7(b) 所示, 可以看出, 当 $T_{60} \leq 0.4$ 时, 本文方法有明显的优势, IVA 基本处于失效的状态。此外, 结合图 4 分析可知, 当信源数和观测信号数增加时 IVA 和 ILRMA 分离性能变得越来越差, 尤其是 IVA 在 $m = n = 4$ 时已经不能满足盲源分离的基本要求。

7 结束语

本文根据频域观测信号的结构特征和全局优化频点排序策略, 研究了一种基于张量分解的 CBSS 方法。首先, 利用观测序列的估计自相关矩阵构造张量数据。然后, 通过张量分解估计出每个频点对应的混合滤波器矩阵达到频点上的频谱分离效果。最后, 采用以功率比为排序测度的全局优化策略以及 MDP 技术同时消除了每个频点上的排序和幅度模糊性并计算出时域分离信号。在虚拟的立体空间中利用声学的图像方法产生 RIR, 并以实测语音作为信源。实验表明, 在不同仿真条件下本文方法的有效性得到了证明。此外, 频域方法的分离性能整体上优于 IVA、ILRMA 和传统 JAD 方法, 尤其是在信源数和观测信号数都增加的情况下。

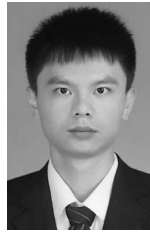
参考文献:

- [1] 张天骐, 马宝泽, 强幸子, 等. 带自适应动量因子的变步长盲源分离方法[J]. 通信学报, 2017, 38(3): 16-24.
ZHANG T Q, MA B Z, QIANG X Z, et al. Variable-step blind source separation method with adaptive momentum factor[J]. Journal on Communications, 2017, 38(3): 16-24.
- [2] MAZUR R, MERTINS A. An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 17(1): 117-126.
- [3] 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.
- [4] KANG F, YANG F R, YANG J. A low-complexity permutation alignment method for frequency-domain blind source separation[J]. Speech Communication, 2019, 115: 88-94.
- [5] KEMIHA M, KACHA A. Complex blind source separation[J]. Circuits,

Systems, and Signal Processing, 2017, 36(11): 4670-4687.

- [6] LEE I, KIM T, LEE T W. Fast fixed-point independent vector analysis algorithms for convolutive blind source separation[J]. Signal Processing, 2007, 87(8): 1859-1871.
- [7] 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.
- [8] FU X, IBRAHIM S, WAI H T, et al. Block-randomized stochastic proximal gradient for low-rank tensor factorization[J]. IEEE Transactions on Signal Processing, 2020, 68: 2170-2185.
- [9] LATHAUWER L D, MOOR B D, VANDEWALLE J. Computation of the canonical decomposition by means of a simultaneous generalized schur decomposition[J]. SIAM Journal on Matrix Analysis and Applications, 2004, 26(2): 295-327.
- [10] YEREDOR A. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation[J]. IEEE Transactions on Signal Processing, 2002, 50(7): 1545-1553.
- [11] LI X L, ADALI T. Complex independent component analysis by entropy bound minimization[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2010, 57(7): 1417-1430.
- [12] 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.
- [13] 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.
- [14] EMURA S, SAWADA H, ARAKI S, et al. Multi-delay sparse approach to residual crosstalk reduction for blind source separation[J]. IEEE Signal Processing Letters, 2020, 27: 1630-1634.

[作者简介]



马宝泽 (1990-), 男, 河北廊坊人, 重庆邮电大学博士生, 主要研究方向为盲信号分离处理、深度学习。



张天骐 (1971-), 男, 四川眉山人, 博士, 重庆邮电大学教授、博士生导师, 主要研究方向为盲信号识别、无线通信的智能信号处理、盲信号处理。



安泽亮 (1993-), 男, 安徽蚌埠人, 重庆邮电大学博士生, 主要研究方向为调制识别、深度学习、盲信号处理。



邓盼 (1990-), 男, 四川宜宾人, 重庆邮电大学博士生, 主要研究方向为信号与信息处理、深度学习。