

文章编号: 1672-2892(2011)06-0765-06

## 弱稀疏语音信号的欠定盲分离

和继威<sup>1</sup>, 刘郁林<sup>1</sup>, 王开<sup>1</sup>, 李力<sup>2</sup>

(1.重庆通信学院 DSP 实验室, 重庆 400035; 2.重庆市教育考试院, 重庆 400035)

**摘要:** 为解决弱稀疏语音信号的欠定盲分离问题, 根据语音信号的部分 W-分离正交性, 提出一种基于单源主导区间的混合矩阵盲估计方法。该方法根据单源主导区间的性质, 通过二元行矢量提取单源观测样本, 对单源观测样本进行 K 均值聚类 and 主成分分析来估计混合矩阵。仿真结果表明, 提出的方法可有效提高分离语音的性能, 与直接利用 K-PCA 方法相比, 分离语音的平均信噪比提高了 10 dB 左右。

**关键词:** 欠定盲分离; 弱稀疏; 单源主导区间; W-分离正交性

**中图分类号:** TN911.25

**文献标识码:** A

## Underdetermined blind separation for weak sparse speech signals

HE Ji-wei<sup>1</sup>, LIU Yu-lin<sup>1</sup>, WANG Kai<sup>1</sup>, LI Li<sup>2</sup>

(1.Chongqing Communication College, Chongqing 400035, China; 2.Chongqing Education Examination College, Chongqing 400035, China)

**Abstract:** In order to solve the problem of underdetermined blind speech separation for weak sparse speech signals, according to the partial approximate W-disjoint orthogonality of weak sparse speech signals, a blind mixing matrix estimation method is proposed based on single-source dominated areas. In this method, according to the property of the single-source dominant areas, single-source observational samples are extracted by a binary row vector; then the mixing matrix is estimated by K means clustering and Principal Component Analysis(PCA) on extracted samples. Simulation results show that the proposed method can effectively improve the performance of separated speech signals, and its averaged Signal To Noise Ratio(SNR) is improved by 10 dB compared to K-PCA method.

**Key words:** underdetermined blind separation; weak sparse; single-source dominant area; W-disjoint orthogonality

语音盲分离问题是近年来信号处理领域和神经网络领域研究的热点之一, 在语音识别和增强、视频会议以及安全监控等诸多领域均有较强的应用需求。目前对语音盲分离的相关研究, 大都要求观测信号的个数不少于声源信号的个数, 而观测信号个数少于声源信号个数的欠定盲分离问题则一直是研究的难点<sup>[1-2]</sup>。近年来, 不少学者利用语音信号在时频域的稀疏性, 根据 W-分离正交性假设<sup>[3]</sup>, 对此问题进行了研究。这些研究大致可分为 2 类: 第 1 类是基于时频二元掩蔽的方法<sup>[4]</sup>, 通过设计二元掩蔽值提取属于同一个源信号的时频点来实现混合语音的分离, 估计出的源信号会丢失一部分能量和信息, 导致分离信号中存在音乐噪声问题; 第 2 类是“两步法”<sup>[5-8]</sup>, 即先估计混合矩阵, 再估计源信号。在欠定情况下, 混合矩阵的估计通常采用直线聚类的方法来完成, 例如 K-PCA 方法<sup>[6]</sup>等。此类方法依赖语音信号的 W-分离正交性假设, 当稀疏性较弱时, 只有部分时频点满足 W-分离正交性假设, 观测信号散点图的直线特征较模糊, 远离直线周边的数据点将严重干扰统计聚类, 使得估计出的混合矩阵出现较大偏差。因此, 本文研究的出发点是解决弱稀疏语音信号的欠定盲分离问题, 提出具有稳健性能的欠定盲分离方法。本文针对弱稀疏语音信号, 根据其部分 W-分离正交性, 提出一种基于单源主导区间的欠定盲分离方法。首先, 根据单源主导区间观测信号的性质, 设计二元行矢量提取位于单源主导区间的观测样本, 剔除非单源观测样本; 其次, 利用 K 均值聚类法对单源观测样本进行分类, 对每类样本进行主成分分析估计出混合矩阵; 最后, 根据估计出的混合矩阵恢复源信号。由于提取出的观测样本满足 W-分离正交性假设, 直线特征清晰, 所以提高了混合矩阵的估计精确度; 而且只对提取出的样本进行聚类, 降低了算法的计算复杂度。

## 1 问题描述

用上标 T 表示矩阵转置, 设  $\mathbf{s}(t)=[s_1(t), s_2(t), \dots, s_n(t)]^T$  为  $n$  个语音源信号,  $\mathbf{x}(t)=[x_1(t), x_2(t), \dots, x_m(t)]^T$  为  $\mathbf{s}(t)$  经过媒介传播后由  $m$  个传感器接收到的信号。  $\mathbf{s}(t)$  和  $\mathbf{x}(t)$  满足方程:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) = \sum_{i=1}^n \mathbf{a}_i s_i(t) + \mathbf{v}(t) \quad (1)$$

式中:  $\mathbf{v}(t)=[v_1(t), v_2(t), \dots, v_m(t)]^T$  表示干扰噪声;  $\mathbf{A}=[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$  是未知的  $m \times n$  混合矩阵 ( $m < n$ ),  $\mathbf{a}_i=[a_{1i}, a_{2i}, \dots, a_{mi}]^T$ 。盲分离的目的就是仅根据观测信号  $\mathbf{x}(t)$ , 在相差一个尺度矩阵和排列矩阵的意义上使得分离信号与源信号的波形保持一致。不考虑噪声的情况下, 每个传感器接收到的信号可表示为:

$$x_i(t) = \sum_{k=1}^n a_{ik} s_k(t), \quad i=1, 2, \dots, m \quad (2)$$

为利用语音信号在时频域的稀疏性, 通过短时傅里叶变换将观测信号转换到时频域:

$$X_i(t, f) = \sum_{r=-L/2}^{L/2-1} x_i(t+r) \text{win}(r) e^{-j2\pi fr} \quad (3)$$

式中:  $f \in \{0, \frac{1}{L}f_s, \dots, \frac{L-1}{L}f_s\}$ ,  $f_s$  为抽样频率;  $\text{win}(r)$  表示两端平滑过渡到零的窗函数;  $L$  为窗长;  $t$  代表时间。

按式(1)进行短时傅里叶变换:

$$\mathbf{X}(t, f) = \mathbf{A}\mathbf{S}(t, f) = \sum_{i=1}^n \mathbf{a}_i S_i(t, f) \quad (4)$$

式中:  $\mathbf{X}(t, f)=[X_1(t, f), X_2(t, f), \dots, X_m(t, f)]^T$  是观测信号的时频域表示;  $\mathbf{S}(t, f)=[S_1(t, f), S_2(t, f), \dots, S_n(t, f)]^T$  为语音源信号的时频域表示。

## 2 语音信号的弱稀疏性

对于  $n$  个语音信号  $s_1(t), s_2(t), \dots, s_n(t)$ , 其时频域为  $S_i(t, f)$  ( $i \in I, I = \{1, 2, \dots, n\}$ )。在时频点  $(t, f)$ , 如果任意两个语音信号之间满足:

$$S_i(t, f) S_k(t, f) = 0 \quad i, k \in I \text{ and } i \neq k \quad (5)$$

则称语音信号  $s_1(t), s_2(t), \dots, s_n(t)$  在时频点  $(t, f)$  满足严格 W-分离正交性。由于语音信号的短时傅里叶变换系数通常不会等于零, 如果某些信号在时频点  $(t, f)$  的时频变换系数接近于零, 式(5)可表示为:

$$S_i(t, f) S_k(t, f) \approx 0 \quad i, k \in I \text{ and } i \neq k \quad (6)$$

此时, 称语音信号  $s_1(t), s_2(t), \dots, s_n(t)$  在时频点  $(t, f)$  满足近似 W-分离正交性<sup>[3]</sup>。

现有的基于稀疏性的欠定语音盲分离方法, 大都假设语音源信号在所有时频点均满足(近似)W-分离正交性假设; 在盲分离中, 这种假设可以理解为在任意时频点至多存在一个语音源信号。实际语音源信号只是在部分时频点满足(近似)W-分离正交性假设; 即在部分时频点可能同时存在多个语音源信号:

$$\begin{cases} |S_i(t, f)| \approx 0, & i \in I_1, I_1 \subseteq I \\ |S_k(t, f)| \neq 0, & k \in I_2, I_2 \subseteq I \\ I_1 \cap I_2 = \emptyset, I_1 \cup I_2 = I, I_2 \text{ contains two elements at least} \end{cases} \quad (7)$$

这种情况称为语音源信号  $s_1(t), s_2(t), \dots, s_n(t)$  满足部分 W-分离正交性。满足部分 W-分离正交性的语音源信号  $s_1(t), s_2(t), \dots, s_n(t)$  具有弱稀疏性。

## 3 基于单源主导区间的混合矩阵盲估计

### 3.1 单源主导区间及其性质

在线性混合系统中, 对于混合信号  $\mathbf{X}(t, f)$  的某一时频点  $(t, f)$ , 设语音源信号  $S_i(t, f)$  的能量为  $Q_i(t, f)$  ( $i=1, 2, \dots, n$ ), 如果存在  $S_i(t, f)$  满足  $Q_i(t, f) \gg \sum_{k=1, k \neq i}^n Q_k(t, f)$  ( $\exists i \in \{1, 2, \dots, n\}$ ), 则称时频点  $(t, f)$  为第  $i$  个源信号的单源主导点。由第  $i$  个源信号的相邻单源主导点构成的区间, 称为第  $i$  个源信号的单源主导区间。显然, 位于单源主导区间的时频点满足近似 W-分离正交性假设。

有效估计混合矩阵, 必须剔除不满足 W-分离正交性的时频点。为便于描述, 用  $\rho$  来表示时频点  $(t, f)$ 。根据

单源主导区间的定义, 如果区间  $[\rho_1, \rho_2]$  为第  $i$  个语音源信号的单源主导区间, 则当  $\forall \rho \in [\rho_1, \rho_2]$  时:

$$\mathbf{X}(\rho) \approx \mathbf{a}_i S_i(\rho) \quad (8)$$

令  $\mathbf{b}_i = [b_{1i}, b_{2i}, \dots, b_{mi}]^T$ , 其中  $b_{li} = |a_{li}| \geq 0$  ( $l=1, 2, \dots, m$ ), 对式(8)取绝对值:

$$|\mathbf{X}(\rho)| \approx |\mathbf{a}_i S_i(\rho)| = \mathbf{b}_i |S_i(\rho)| \quad (9)$$

对  $|\mathbf{X}(\rho)|$  进行归一化:

$$\mathbf{u}(\rho) = \frac{|\mathbf{X}(\rho)|}{\|\mathbf{X}(\rho)\|} \approx \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|} \quad (10)$$

令  $\mathbf{U} = [\mathbf{u}(1), \mathbf{u}(2), \dots, \mathbf{u}(K)]$ ,  $K$  表示总的时频点个数。对  $\mathbf{u}(\rho)$  和  $\mathbf{u}(\rho+1)$  作差分:

$$\mathbf{g}(\rho) = \mathbf{u}(\rho) - \mathbf{u}(\rho+1) \approx \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|} - \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|} = \mathbf{0} \quad (11)$$

显然, 如果区间  $[\rho_1, \rho_2]$  为单源主导区间, 则可以找到一个接近于 0 的非负实数  $\varepsilon$ , 使得  $\forall \rho \in [\rho_1, \rho_2]$  都有  $\|\mathbf{g}(\rho)\|_1 \leq \varepsilon$ 。

### 3.2 单源主导区间的提取

令  $\mathbf{X} = [\mathbf{X}(1), \mathbf{X}(2), \dots, \mathbf{X}(K)]$  表示观测信号矩阵,  $\mathbf{G} = [\mathbf{g}(1), \mathbf{g}(2), \dots, \mathbf{g}(K)]$  表示对矩阵  $\mathbf{U}$  的相邻两个列矢量作差分后的矩阵。为提取单源主导区间, 设计 1 个  $1 \times K$  的二元行矢量  $\mathbf{F} = [\mathbf{f}(1), \mathbf{f}(2), \dots, \mathbf{f}(K)]$ , 其中:

$$\mathbf{f}(\rho) = \begin{cases} 1, & \rho \in \text{single-source dominant area} \\ 0, & \text{else} \end{cases}, \quad \rho = 1, 2, \dots, K \quad (12)$$

当矩阵  $\mathbf{F}$  与矩阵  $\mathbf{X}$  每 1 行的对应元素相乘时, 单源主导区间的数据保持不变, 非单源区间的数据将变为零, 这样就可以提取出单源主导区间的观测样本。

### 3.3 混合矩阵的估计

将位于单源主导区间的观测样本记为  $\mathbf{Y} = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(L)\}$ ,  $\mathbf{y}(i)$  表示单源主导区间的第  $i$  个样本,  $L$  表示总的样本数。对  $\mathbf{Y}$  中的数据样本进行标准化处理:

$$\mathbf{y}'(i) = \begin{cases} \frac{\mathbf{y}(i)}{\|\mathbf{y}(i)\|}, & y_1(i) \geq 0, \\ -\frac{\mathbf{y}(i)}{\|\mathbf{y}(i)\|}, & y_1(i) < 0, \end{cases} \quad i = 1, 2, \dots, L \quad (13)$$

式中  $y_1(i)$  表示  $\mathbf{y}(i)$  的第 1 个分量。记标准化后的数据为  $\mathbf{Y}' = \{\mathbf{y}'(1), \mathbf{y}'(2), \dots, \mathbf{y}'(L)\}$ 。对  $\mathbf{Y}'$  中的数据进行  $K$  均值聚类, 将其聚为  $n$  类, 对应的  $\mathbf{Y}$  中的数据被分成  $n$  类, 依次记为  $y^1, y^2, \dots, y^n$ 。分别对  $y^i, i=1, 2, \dots, n$  进行主成分分析, 对应最大特征值的  $n$  个特征向量, 即是混合矩阵各个列向量的估计。

## 4 分离模型及源信号的估计

对于欠定模型, 稀疏信号盲分离归结为求解如下优化问题<sup>[6,9]</sup>:

$$\min_{\mathbf{A}, \mathbf{S}} \frac{1}{2\sigma^2} \|\mathbf{AS} - \mathbf{X}\|^2 + \sum_{i, \rho} |S_i(\rho)| \quad (14)$$

式中:  $\sigma^2$  表示噪声方差; 第 1 项为重构误差平方和, 第 2 项为非稀疏惩罚项(假设源信号相互独立, 服从 Laplace 分布)。式(14)模型是一个多变量优化问题, 直接求解比较困难。如果  $\mathbf{A}$  已经事先给定, 则模型可简化为:

$$\min_{\mathbf{S}(\rho)} \frac{1}{2\sigma^2} \|\mathbf{AS}(\rho) - \mathbf{X}(\rho)\|^2 + \sum_{i=1}^n |S_i(\rho)|, \quad \rho = 1, 2, \dots, K \quad (15)$$

在不考虑噪声的情况下, 上式退化为:

$$\min_{\mathbf{S}(\rho)} \sum_{i=1}^n |S_i(\rho)| \quad \text{s.t. } \mathbf{AS}(\rho) = \mathbf{X}(\rho), \quad \rho = 1, 2, \dots, K \quad (16)$$

在给定混合矩阵  $\mathbf{A}$  的基础上, 每个时频点  $\rho$  确定一个优化问题, 从而式(16)的优化问题可分解为  $K$  个容易处理的小问题。根据估计出的混合矩阵  $\hat{\mathbf{A}} = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_n]$ , 求出由其中  $m$  个列组成的子矩阵  $\mathbf{B}_r = [\hat{\mathbf{a}}_{r1}, \hat{\mathbf{a}}_{r2}, \dots, \hat{\mathbf{a}}_{rm}]$  的逆, 其中  $\{r1, r2, \dots, rm\} \subseteq \{1, 2, \dots, n\}$ ,  $r = 1, 2, \dots, C_n^m$ , 计算

$$\begin{cases} \mathbf{S}^{(r)}(\rho) = \mathbf{B}_r^{-1} \mathbf{X}(\rho), & r=1,2,\dots,C_n^m \\ \mathbf{S}^{(p)}(\rho) = 0, & p \neq r_1, r_2, \dots, r_m \end{cases} \quad (17)$$

在时频点  $\rho$ , 如果对任意的  $r, q \in \{1, 2, \dots, C_n^m\}$   $r \neq q$ , 存在  $\sum |\mathbf{S}^{(r)}(\rho)| \leq \sum |\mathbf{S}^{(q)}(\rho)|$ , 则可通过式(17)求得  $\hat{\mathbf{S}}(\rho)$ 。求出源信号在所有时频点的估计值, 对其进行短时傅里叶逆变换, 便可得到源信号的估计。

## 5 实验仿真

本文分别对  $m=2, n=4$  和  $m=3, n=4$  的欠定混合情况进行了实验仿真。4 个满足部分 W-分离正交性的弱稀疏语音源信号的时长均为 2.5 s, 抽样频率为 8 kHz。采用的混合矩阵分别为:

$$\mathbf{A}_{2 \times 4} = \begin{bmatrix} 0.9995 & 0.8823 & -0.2491 & 0.4238 \\ -0.0316 & 0.4708 & 0.9685 & 0.9058 \end{bmatrix}, \quad \mathbf{A}_{3 \times 4} = \begin{bmatrix} 0.9586 & 0.8014 & -0.1502 & 0.2199 \\ -0.0321 & 0.4697 & 0.9532 & 0.9062 \\ 0.2803 & 0.3692 & 0.2003 & 0.3605 \end{bmatrix}$$

对观测信号进行短时傅里叶变换时采用 Hamming 窗, 窗长 512 点, 帧移 256 点。变换后观测信号的散点图分别如图 1(a)、图 2(a)所示。提取位于单源主导区间的观测样本时, 误差  $\varepsilon=0.005$ , 提取出的位于单源主导区间的观测样本的散点图分别如图 1(b)、图 2(b)所示。提取出的观测样本满足近似 W-分离正交性假设, 具有清晰的直线特征, 利用其估计混合矩阵, 能够提高混合矩阵的估计精确度。

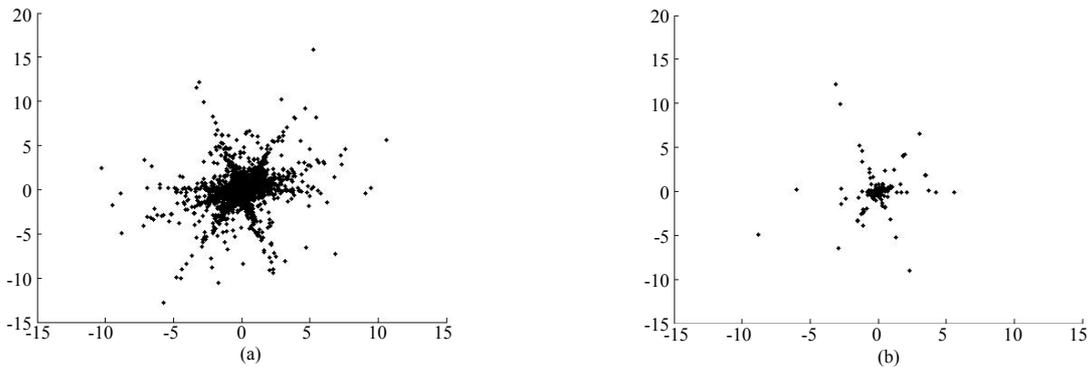


Fig.1 Scatter plots for  $2 \times 4$  mixing case  
图 1  $2 \times 4$  混合时的散点图

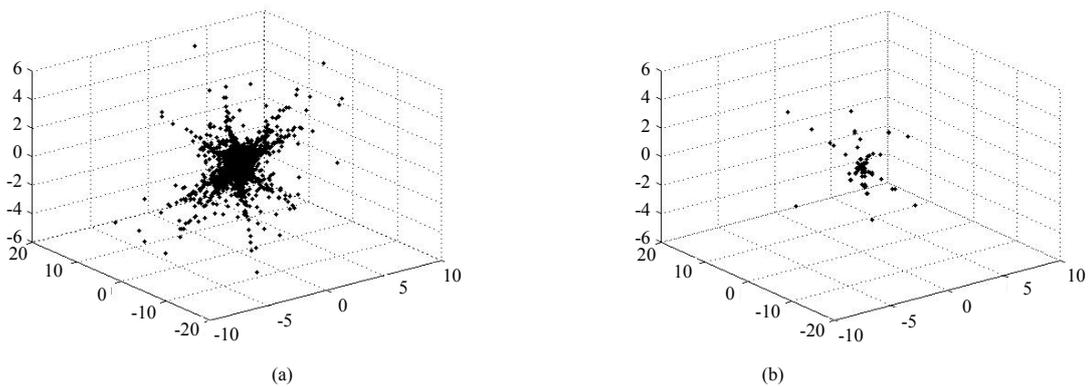


Fig.2 Scatter plots for  $3 \times 4$  mixing case  
图 2  $3 \times 4$  混合时的散点图

为了定量分析本文所给方法的有效性, 从两个方面评判盲分离效果。一方面, 采用广义串扰误差比较混合矩阵的估计性能, 其定义如下:

$$GCE = \min_{\mathbf{M} \in \Pi} \|\mathbf{A} - \hat{\mathbf{A}}\mathbf{M}\| \quad (18)$$

式中:  $\Pi$  表示  $n \times n$  维可逆阵的集合, 且每 1 列只有 1 个非零元素;  $\|\cdot\|$  表示单位范数。当且仅当  $\mathbf{A}$  和  $\hat{\mathbf{A}}$  为等效矩阵时,  $GCE=0$ ; 当  $GCE \leq 0.1$  时, 估计性能已足够好。另一方面, 比较分离信号  $\hat{s}$  与对应语音源信号  $s$  之间的信噪比, 其定义如下:

$$R_{SN,i} = 10 \lg \left( \frac{\sum_{t=1}^T s_i^2(t)}{\sum_{t=1}^T (s_i(t) - \hat{s}_i(t))^2} \right) \text{ (dB)} \quad (19)$$

$R_{SN,i}$  越大, 分离效果越好。

首先比较本文方法和 K-PCA 方法的混合矩阵估计性能, 仿真结果如表 1 所示。由于只有部分时频点满足近似 W-分离正交性假设, 所以 K-PCA 方法的估计精确度不高。本文方法利用单源主导区间的观测样本估计混合矩阵, 由于位于单源主导区间的观测样本满足近似 W-分离正交性假设, 所以混合矩阵的估计性能得到了提高。

其次比较不同混合情况下算法的分离性能,  $2 \times 4$  混合时分离性能如表 2 所示,  $3 \times 4$  混合时分离性能如表 3 所示。在“两步法”中, 混合矩阵估计精确度的高低决定了分离性能的优劣。由仿真结果可知, 本文方法的分离性能远优于 K-PCA 方法。利用本文方法估计出来的混合矩阵和原始混合矩阵恢复语音源信号, 二者的分离性能非常接近, 这也从另一个方面说明了利用本文方法估计混合矩阵的精确度非常高。

## 6 结论

本文采用先估计混合矩阵再估计源信号的“两步法”, 对弱稀疏语音信号的欠定盲分离问题进行了研究。为了能更好地利用语音信号的稀疏性, 提高混合矩阵的估计精确度, 本文在分析单源主导区间观测信号性质的基础上, 提出了一种基于单源主导区间的混合矩阵盲估计方法。该方法计算复杂度低, 估计精确度高, 具有较好的鲁棒性。下一步将研究其在欠定卷积混合语音盲分离中的应用。

### 参考文献:

- [1] Choi S, Cichocki A. Blind source separation and independent component analysis: a review[J]. *Neural Information Processing*, 2005, 6(1):1-57.
- [2] 李木森, 毛剑琴. 盲信号分离的现状和展望[J]. *信息与电子工程*, 2003, 1(1):69-79. (LI Musen, MAO Jianqin. A survey of blind source separation[J]. *Information and electronic engineering*, 2003, 1(1):69-79.)
- [3] Rickard S, Yilmaz Z. On the approximate W-disjoint orthogonality of speech[C]// *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech and Signal Processing: ICASSP'02*. Washington, DC: IEEE Computer Society, 2002:529-532.
- [4] Yilmaz O, Rickard S. Blind separation of speech mixtures via time-frequency masking[J]. *IEEE Trans. on Signal Processing*, 2004, 52(7):1830-1847.
- [5] Bofill P, Zibulevsky M. Underdetermined blind source separation using sparse representations[J]. *Signal Processing*, 2001, 81(11):2353-2362.
- [6] 何昭水, 谢胜利, 傅予力. 稀疏表示与病态混叠盲分离[J]. *中国科学 E 辑: 信息科学*, 2006, 36(8):864-879. (HE Zhaoshui, XIE Shengli, FU Yuli. Sparse representation and blind source separation of ill-posed mixtures[J]. *Science in China Series E: Information Science*, 2006, 36(8):864-879.)
- [7] Li Y, Amari A, Cichocki A. Underdetermined blind source separation based on sparse representation [J]. *IEEE Trans. on Signal Processing*, 2006, 54(2):423-437.
- [8] Wei Z, Ju L, Jiande S. A new two-step approach to underdetermined blind source separation using sparse representation[C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007:953-956.
- [9] 谭北海, 谢胜利. 基于源信号数目估计的欠定盲分离[J]. *电子与信息学报*, 2008, 30(4):863-867. (TAN Beihai, XIE Shengli. Underdetermined blind separation based on source signals' number estimation[J]. *Journal of electronics and information technology*, 2008, 30(4):863-867.)

(下转第 781 页)

表 1 混合矩阵估计性能比较

mixed mode	K-PCA method	proposed method
2×4	0.496 4	0.015 2
3×4	0.438 7	0.012 9

表 2 2×4 混合时分离性能比较

estimated method of mixing matrix	$R_{SN,1}$	$R_{SN,2}$	$R_{SN,3}$	$R_{SN,4}$	mean
K-PCA method	16.321 4	14.543 8	13.768 6	17.562 2	15.549 0
proposed method	26.789 4	25.302 1	23.552 3	27.298 6	25.735 6
original matrix	26.867 5	25.324 7	23.963 4	27.347 6	25.875 8

表 3 3×4 混合时分离性能比较

estimated method of mixing matrix	$R_{SN,1}$	$R_{SN,2}$	$R_{SN,3}$	$R_{SN,4}$	mean
K-PCA method	21.158 6	20.541 5	18.103 7	21.980 6	20.446 1
proposed method	32.453 6	30.214 7	29.473 1	33.271 4	31.353 2
original matrix	32.547 5	30.302 1	29.513 4	33.314 2	31.419 3