

弱稀疏性下的欠定语音盲分离方法

王国鹏, 刘郁林, 罗颖光

(重庆通信学院 DSP 研究室, 重庆 400035)

摘要: 针对语音信号的弱稀疏性, 提出一种新的基于混合矩阵估计的欠定语音盲分离方法。该方法通过主成分分析检测只有一个源信号存在时的时频点并用于估计混合矩阵, 从而克服语音信号稀疏性变弱时的影响, 提高混合矩阵估计精度。结合子空间法重构源信号, 进一步提高分离性能, 并从几何角度证明子空间方法, 仿真结果表明该方法的分离性能优于 Cluster-UBSS, 且鲁棒性更好。

关键词: 语音盲分离; 混合矩阵估计; 稀疏性; 主成分分析; 子空间方法

Underdetermined Blind Speech Separation Method Under Weak Sparseness

WANG Guo-peng, LIU Yu-lin, LUO Ying-guang

(DSP Lab, Chongqing Communication College, Chongqing 400035)

【Abstract】 This paper proposes a new method based on mixing matrix estimation for underdetermined blind speech separation, aiming at speech signals under weak sparseness. The method detects and exploits time-frequency bins with only one source by Principal Component Analysis (PCA) to estimate mixing matrix, it overcomes the shortcoming of weak sparseness of speech signals and improves the estimation precision of mixing matrix. It combines a subspace method to reconstruct the sources to improve the separation performance further. The subspace method is proved from geometric interpretation. Simulation results show the separability of the method is better than Cluster-BSS, and its robustness is better.

【Key words】 blind speech separation; mixing matrix estimation; sparseness; Principal Component Analysis(PCA); subspace method

1 概述

盲源分离指由观测信号分离各源信号的方法^[1], 其应用广泛, 在语音、通信和生物医学工程等多个领域具有重要的理论价值和实际意义。

本文主要研究欠定混合下的语音盲分离(源信号数目大于传感器数目), 此时无法估计混合矩阵, 常用的独立分量分析方法不适用。

基于稀疏性的盲分离方法^[2-3]越来越受到关注, 较有代表性的是时频二元掩蔽^[2], 该方法利用语音信号的 W-分离正交性假设(W-DO), 但 W-DO 只允许每个时频点至多存在 1 个源信号, 会导致音乐噪声。

文献[3]提出一种基于混合矩阵估计的盲分离方法 Cluster-UBSS, 通过聚类估计混合矩阵, 结合子空间方法重构源信号, 该方法允许每个时频点存在多个源信号(但要少于传感器数目), 能有效减少音乐噪声, 但在估计混合矩阵时仍基于 W-DO 假设, 因此, 估计精度不高, 在弱稀疏性下估计性能下降很快。

针对弱稀疏性, 本文提出一种新的基于混合矩阵估计的欠定语音盲分离方法。利用主成分分析(Principal Component Analysis, PCA)检测只有一个源信号存在的时频点, 此类时频点的最大特征值对应的特征向量即为一个混合向量估计。对所有估计的混合向量进行 K 均值聚类, 将聚类中心作为混合向量的最终估计, 从而估计出混合矩阵。

2 问题描述

盲源分离的实时线性混合模型为

$$\mathbf{x}(t) = \mathbf{A}s(t) \quad (1)$$

其中, $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ 表示混合信号向量; $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ 表示源信号向量; $\mathbf{A} = [a_1, a_2, \dots, a_N]$ 为 $M \times N$ 的混合矩阵, M 表示传感器数目, N 表示源信号数目; a_i 为混合向量。本文只考虑欠定混合情况 ($M < N$)。

3 语音信号的弱稀疏性

语音信号在时频域具有稀疏性, 因此, 通过短时傅里叶变换(STFT)将信号转换为时频域表示:

$$X(\omega, t) = \mathbf{A}S(\omega, t) \quad (2)$$

设 $S_i(\omega, t)$ 和 $S_j(\omega, t)$ 是任意 2 个语音信号的时频表示, W-DO 描述为^[2]

$$S_i(\omega, t)S_j(\omega, t) = 0, \quad \forall \omega, t \quad (3)$$

每个时频点上至多存在 1 个源信号。

$M=3, N=4$ 时的混合信号散点如图 1 所示。

若语音信号满足 W-DO, 则稀疏性很强, 散点图应呈现出清晰的直线方向, 如图 1(a)所示。

实际语音信号并不能完全满足 W-DO, 单个时频点可能存在多个源信号, 此时散点图呈现的直线方向并不清晰, 如图 1(b)所示, 本文称此情况为弱稀疏性。

基金项目: 国家自然科学基金资助项目(60672157, 60672158); 重庆市自然科学基金资助项目(CSTC2005BB4219)

作者简介: 王国鹏(1983 -), 男, 硕士研究生, 主研方向: 盲源分离; 刘郁林, 教授、博士; 罗颖光, 硕士研究生

收稿日期: 2009-03-22 **E-mail:** wgphebei@126.com

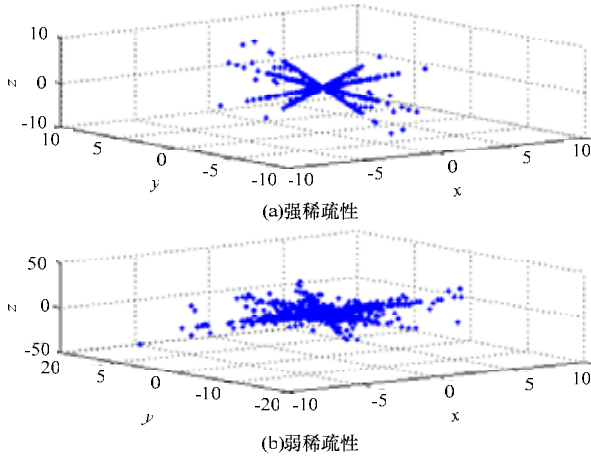


图1 $M=3, N=4$ 时的混合信号散点

4 基于PCA的混合矩阵估计

4.1 混合信号的PCA

假设混合信号向量 $X(\omega, t)$ 的均值为 0，对其协方差矩阵进行特征值分解：

$$R_X(\omega, t) = E[X(\omega, t)X(\omega, t)^H] = AR_S(\omega, t)A^H = UDU^H \quad (4)$$

其中， $U = [u_1, u_2, \dots, u_M]$ 为特征向量矩阵； $D = \text{diag}[\sigma_1^2(\omega, t), \sigma_2^2(\omega, t), \dots, \sigma_M^2(\omega, t)]$ 为特征值矩阵； $(\cdot)^H$ 表示共轭转置。假设该时频点满足 W-DO 且只有第 m 个源信号存在，其他信号的功率均为 0：

$$\sigma_m^2(\omega, t) \neq 0, \sigma_i^2(\omega, t) = 0, (i \neq m) \quad (5)$$

称此时频点为单一源信号时频点，式(4)简化为

$$R_X(\omega, t) = u_m \sigma_m^2(\omega, t) u_m^H \quad (6)$$

式(2)简化为

$$X(\omega, t) = a_m S_m(\omega, t) \quad (7)$$

其协方差矩阵：

$$R_X(\omega, t) = a_m E[S_m(\omega, t)S_m(\omega, t)^H] a_m^H \quad (8)$$

比较式(6)和式(8)可以看出， u_m 是混合向量 a_m 的一个估计，两者呈比例关系。

4.2 单一源信号时频点的检测

考虑到噪声的影响，式(4)的特征值满足：

$$\sigma_m^2(\omega) \gg \sigma_i^2(\omega), (i=1, 2, \dots, M, i \neq m) \quad (9)$$

假设特征值 $\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2$ ，引入一个阈值：

$$\varepsilon = 1 - \frac{\sqrt{\sigma_2^2}}{\sqrt{\sigma_1^2}} \quad (10)$$

当 ε 很大时，说明只有 1 个源信号的能量非常大，检测出该时频点只有 1 个源信号。

4.3 混合矩阵的估计

所有检测的单一源信号时频点的最大特征值对应的特征向量构成 N 个混合向量估计的集合，通过 K 均值聚类将估计结果聚为 N 类，将聚类中心作为混合向量的最终估计，从而估计出混合矩阵。

5 基于子空间的源信号重构

5.1 基本原理

假设每个时频点存在的源信号数目少于传感器数目，式(2)退化为

$$X(\omega, t) = a_{\alpha_1} S_{\alpha_1}(\omega, t) + a_{\alpha_2} S_{\alpha_2}(\omega, t) + \dots + a_{\alpha_K} S_{\alpha_K}(\omega, t) \quad (11)$$

其中， $K \leq M-1$ ； $[a_{\alpha_1}, a_{\alpha_2}, \dots, a_{\alpha_K}]$ 是 $[a_1, a_2, \dots, a_N]$ 的一个子集。式(11)表明混合信号位于部分基向量张成的线性子空间内，令

$$\tilde{A} = [a_{\alpha_1}, a_{\alpha_2}, \dots, a_{\alpha_K}] \quad (12)$$

将 \tilde{A} 称为信号子空间。设 $Q_{\tilde{A}}$ 为 \tilde{A} 对应的噪声子空间的正交投影矩阵：

$$Q_{\tilde{A}} = I - \tilde{A}(\tilde{A}^H \tilde{A})^{-1} \tilde{A}^H \quad (13)$$

其中， I 为单位矩阵。根据 $Q_{\tilde{A}}$ 的性质得：

$$\begin{cases} Q_{\tilde{A}} a_i = 0, & i \in \{\alpha_1, \alpha_2, \dots, \alpha_K\} \\ Q_{\tilde{A}} a_i \neq 0, & \text{其他} \end{cases} \quad (14)$$

在 A 已知的情况下确定张成子空间 \tilde{A} 的基向量 $(a_{\alpha_1}, a_{\alpha_2}, \dots, a_{\alpha_K})$ ：

$$\{\alpha_1, \alpha_2, \dots, \alpha_K\} = \min_{\beta_1, \beta_2, \dots, \beta_K} \left\{ \|Q_{\tilde{A}_\beta} X(\omega, t)\| \mid \tilde{A}_{\beta} \right\} \quad (15)$$

其中， $\tilde{A}_{\beta} = [a_{\beta_1}, a_{\beta_2}, \dots, a_{\beta_K}]$ ， $K \in \{1, 2, \dots, M-1\}$ 。重构信号为

$$y(\omega, t) = \tilde{A}^{\#} X(\omega, t) \quad (16)$$

其中， $(\cdot)^{\#}$ 表示伪逆。对 $y(\omega, t)$ 进行短时傅里叶逆变换得到时域分离信号 $y(t)$ 。

5.2 子空间方法证明

将 $X(\omega, t)$ 看作是 M 维空间中的一个点，求 \tilde{A} 即求 $X(\omega, t)$ 所在的子空间。理想情况下， $X(\omega, t)$ 到其所在子空间的距离为 0，但考虑噪声的影响， $X(\omega, t)$ 不可能恰好位于子空间中，因此，求 \tilde{A} 的问题转化为在所有子空间中，求与 $X(\omega, t)$ 距离最近的子空间。空间点到子空间的距离可以理解为在子空间中寻找一个点，使 2 个点的距离最小，点到子空间的距离如图 2 所示。

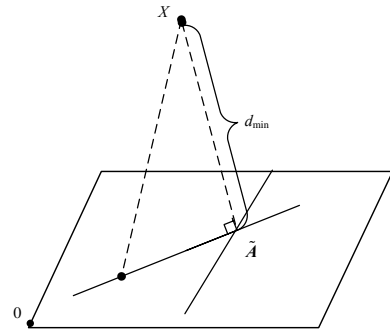


图2 点到子空间的距离

将 $X(\omega, t)$ 记为 X ，定义点到子空间的最小距离^[4]：

$$d_{\min} = \sqrt{X^H X - X^H \tilde{A}(\tilde{A}^H \tilde{A})^{-1} \tilde{A}^H X} \quad (17)$$

将 $X(\omega, t)$ 到所有子空间的最小距离进行比较，最小值对应的子空间即为 $X(\omega, t)$ 所在的子空间。即

$$\{\alpha_1, \alpha_2, \dots, \alpha_K\} = \arg \min_{\beta_1, \beta_2, \dots, \beta_K} \left\{ d_{\min} \mid \tilde{A}_{\beta} \right\} \quad (18)$$

具体证明步骤如下：

式(17)可以化为

$$d_{\min} = \sqrt{X^H X - X^H (I + \tilde{A}(\tilde{A}^H \tilde{A})^{-1} \tilde{A}^H - I) X} \quad (19)$$

根据式(19)和式(13)得：

$$d_{\min} = \sqrt{X^H X - X^H X + X^H (I - \tilde{A}(\tilde{A}^H \tilde{A})^{-1} \tilde{A}^H) X} = \sqrt{X^H Q_{\tilde{A}} X} \quad (20)$$

由于正交投影矩阵 \mathbf{Q} 是一个幂等矩阵和 Hermitian 矩阵, 因此 $\mathbf{Q}^2 = \mathbf{Q}$ 和 $\mathbf{Q}^H = \mathbf{Q}$, 即

$$d_{\min} = \sqrt{\mathbf{X}^H \mathbf{Q} \mathbf{Q} \mathbf{X}} = \sqrt{\mathbf{X}^H \mathbf{Q}^H \mathbf{Q} \mathbf{X}} = \sqrt{(\mathbf{Q} \mathbf{X})^H \mathbf{Q} \mathbf{X}} = \sqrt{\|\mathbf{Q} \mathbf{X}\|^2} = \|\mathbf{Q} \mathbf{X}\| \quad (21)$$

将式(21)代入式(18)得:

$$\{\alpha_1, \alpha_2, \dots, \alpha_K\} = \arg \min_{\beta_1, \beta_2, \dots, \beta_K} \left\{ \left\| \mathbf{Q}_{\beta_j} \mathbf{X}(\omega, t) \right\| \left\| \tilde{\mathbf{A}}_{\beta_j} \right\| \right\} \quad (22)$$

由此可知, 式(22)和式(15)是等价的。

6 仿真实验分析

本文分别对 $M=2, N=3$ (3 个女声) 以及 $M=3, N=4$ (3 个女声、1 个男声) 的混合情况进行仿真。语音长 5 s, 抽样频率 10 kHz, STFT 采用 Hamming 窗, 窗长 512 点, 帧移 256 点, $\varepsilon=0.99$ 。

定义 1 混合矩阵估计的性能评价指标, 表述原矩阵和估计矩阵之间的广义干扰误差^[5](Generalized Crosstalking Error, GCE):

$$GCE = \min_{M \in \mathcal{M}} \left\| \mathbf{A} - \hat{\mathbf{A}} \mathbf{M} \right\| \quad (23)$$

其中, \mathbf{M} 是 $N \times N$ 可逆矩阵, 每列仅有一个非零元素; $\|\cdot\|$ 表示单位范数; $\hat{\mathbf{A}}$ 是 \mathbf{A} 的估计, 两者相等时 GCE 为 0。

定义 2 分离性能评价指标, 表示为信干比^[2](Signal to Interference Ratio, SIR):

$$SIR_i = 10 \lg \left(\frac{\sum_t y_i^s(t)^2}{\sum_t y_i^{if}(t)^2} \right) \quad (24)$$

其中, $y_i^s(t)$ 是分离信号 $y_i(t)$ 中来自源信号 $s_i(t)$ 的成分; $y_i^{if}(t)$ 是来自于其他源信号的干扰成分; SIR_i 越大分离效果越好, SIR_i 单位为 dB。

仿真采用的混合矩阵为

$$\mathbf{A}_{2 \times 3} = \begin{bmatrix} 0.923 & 9 & 0.809 & 0 & 0.500 & 0 \\ 0.382 & 7 & 0.587 & 8 & 0.866 & 0 \end{bmatrix}$$

$$\mathbf{A}_{3 \times 4} = \begin{bmatrix} 0.080 & 5 & 0.168 & 2 & 0.171 & 0 & 0.415 & 6 \\ 0.033 & 4 & 0.122 & 2 & 0.296 & 2 & 0.415 & 6 \\ 0.996 & 2 & 0.978 & 1 & 0.939 & 7 & 0.809 & 0 \end{bmatrix}$$

将本文提出方法与 Cluster-UBSS 方法进行比较。

当比较混合矩阵估计性能时, 本文方法的估计性能远优于 Cluster-UBSS。Cluster-UBSS 在估计混合矩阵时采用 W-DO 假设, 其估计精度依赖于信号的稀疏性, 当稀疏性很好时估计性能很好, 反之下降, 因此鲁棒性较差。Cluster-UBSS 在 2×3 混合情况下的估计性能要好于 3×4 的混合情况, 混合情况越简单, 稀疏性越强, 估计性能也越好。本文方法减小了对稀疏性的依赖, 通过检测满足 W-DO 的局部时频点来估计混合矩阵。因此, 在稀疏性较差时, 估计性能变化不大, 具有很好的鲁棒性。混合矩阵估计性能比较如表 1 所示。

表 1 混合矩阵估计性能比较

混合方式	Cluster-UBSS	本文方法
2×3	$\hat{\mathbf{A}}_{2 \times 3} = \begin{bmatrix} 0.922 & 7 & 0.809 & 9 & 0.498 & 2 \\ 0.385 & 1 & 0.586 & 4 & 0.867 & 1 \end{bmatrix}$ GCE = 0.291 3	$\hat{\mathbf{A}}_{2 \times 3} = \begin{bmatrix} 0.9227 & 0.8099 & 0.4982 \\ 0.3851 & 0.5864 & 0.8671 \end{bmatrix}$ GCE = 0.008 7
3×4	$\hat{\mathbf{A}}_{3 \times 4} = \begin{bmatrix} 0.110 & 0 & 0.109 & 2 & 0.427 & 6 & 0.311 & 8 \\ 0.073 & 5 & 0.165 & 2 & 0.436 & 5 & 0.320 & 3 \\ 0.991 & 2 & 0.980 & 2 & 0.791 & 6 & 0.894 & 6 \end{bmatrix}$ GCE = 0.438 7	$\hat{\mathbf{A}}_{3 \times 4} = \begin{bmatrix} 0.080 & 3 & 0.174 & 3 & 0.185 & 5 & 0.415 & 6 \\ 0.033 & 7 & 0.130 & 5 & 0.290 & 4 & 0.415 & 4 \\ 0.996 & 2 & 0.975 & 7 & 0.937 & 8 & 0.809 & 1 \end{bmatrix}$ GCE = 0.038 9

比较不同混合情况下 2 种算法的分离性能, 2×3 混合时分离性能比较如表 2 所示, 3×4 混合时分离性能比较如表 3 所示。

表 2 2×3 混合时分离性能比较

	SIR_1	SIR_2	SIR_3	平均值
Cluster-UBSS	21.78	26.29	52.27	33.45
本文方法	42.34	28.14	47.08	39.19

表 3 3×4 混合时分离性能比较

	SIR_1	SIR_2	SIR_3	SIR_4	平均值
Cluster-UBSS	0.84	35.05	13.66	-16.01	8.39
本文方法	32.04	30.08	44.84	31.36	34.58

由于子空间方法与混合矩阵有关, 因此混合矩阵估计精度的高低决定了分离性能的优劣。从仿真结果可知, 本文方法的混合矩阵估计性能比 Cluster-UBSS 好, 其分离性能得到相应提高, 在实际试听中, 音乐噪声也减少。

7 结束语

本文针对语音信号的弱稀疏性, 提出一种基于 PCA 的混合矩阵估计方法, 该方法具有较好的鲁棒性。下一步将研究其与稀疏性的具体关系, 以及在接近实际情况的卷积混合中的应用。

参考文献

- [1] Haykin S. Unsupervised Adaptive Filtering: Blind Source Separation[M]. New York, USA: Wiley J & Sons, 2000.
- [2] Rickard S, Yilmaz Z. On the Approximate W-disjoint Orthogonality of Speech[C]//Proc. of ICASSP'02. Orlando, Florida, USA: [s. n.], 2002: 529-532.
- [3] Aissa E B A, Meraim A K, Grenier Y. Underdetermined Blind Source Separation of Audio Sources in Time-frequency Domain[C]//Proc. of the Signal Processing with Adaptive Sparse Structured Representations. Rennes, France: [s. n.], 2005.
- [4] He Z S, Cichocki A. K-subspace Clustering and Its Application in Sparse Component Analysis[C]//Proceedings of the 14th European Symposium on Artificial Neural Networks. Bruges, Belgium: [s. n.], 2006.
- [5] Theis F J, Langa E W, Puntonet C G. A Geometric Algorithm for Overcomplete Linear ICA[J]. Neuro Computing, 2004, 56(18): 381-398.

编辑 陆燕菲

(上接第 245 页)

参考文献

- [1] 曹小平, 路广安. 装备维修器材保障[M]. 北京: 国防大学出版社, 2005.
- [2] 曹继平, 宋建社. 一种战时装备维修保障资源优化调度算法[J]. 系统仿真学报, 2007, 19(15): 3390-3394.
- [3] 徐航, 张会奇, 陈春良. 基于遗传算法的战时维修保障系统优

化研究[J]. 装甲兵工程学院学报, 2006, 20(5): 1-4.

- [4] 朱昱, 宋建社, 王正元. 一种基于最大保障时间的战时装备维修任务调度[J]. 系统工程与电子技术, 2007, 29(11): 1900-1903.
- [5] 陈冲, 张莹. 基于 Extend 的大型露天矿卡车调度系统仿真[J]. 系统仿真学报, 2007, 19(4): 914-917.

编辑 张帆