

文章编号:1001-9081(2009)04-1056-03

基于稀疏性的欠定语音盲分离方法研究

王国鹏, 刘郁林, 罗颖光

(重庆通信学院 DSP 研究室, 重庆 400035)

(wgphebei@126.com)

摘要: 针对源信号增多导致语音信号稀疏性变差的问题, 提出一种新的基于稀疏性的混合矩阵估计方法, 利用主分量分析(PCA)检测只有一个源信号存在的时频点并用于估计混合矩阵, 从而提高了估计性能, 特别适用于欠定语音盲分离。同时指出了影响基于稀疏性语音盲分离方法性能的因素。仿真结果验证了上述结论。

关键词: 稀疏性; 混合矩阵估计; 语音盲分离

中图分类号: TP912 **文献标志码:** A

Underdetermined blind speech separation of sparseness

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

(DSP Laboratory, Chongqing Communication College, Chongqing 400035, China)

Abstract: A new sparseness-based method was proposed for mixing matrix estimation, in the case of poor sparseness of speech signals with increasing number of sources. The time-frequency bins with only one source were detected by Principal Component Analysis (PCA), and then were exploited to estimate the mixing matrix to improve the estimation performance. The proposed method is especially applicable to underdetermined blind speech separation. The reasons deteriorating the performance of blind speech separation were also pointed out. The simulation results demonstrate the conclusions above.

Key words: sparseness; mixing matrix estimation; blind speech separation

0 引言

语音信号盲分离^[1]是信号处理领域的一个活跃分支, 也是盲源分离技术应用的一个研究热点。目前语音盲分离面临的一个挑战是源信号多于传感器的混合情况, 即欠定混合。在这种情况下, 混合矩阵的逆不存在, 传统的独立分量分析方法不再适用, 因此必须寻找一种新方法来解决欠定混合问题。

基于稀疏性的欠定盲分离方法得到了广泛研究。对于语音信号来讲, 稀疏性是指近似 W-分离正交性, 即存在部分时频点, 每一个时频点上至多存在一个源信号, 称此类时频点满足 W-分离正交^{[2]529}。基于稀疏性的语音盲分离方法大致分为两类, 第一类方法首先估计混合矩阵, 然后再估计源信号^[3]; 另一类方法利用时频二元掩蔽^[4]来提取源信号, 省略了对混合矩阵的估计, 但由于估计出的源信号丢失了一部分能量和信息, 其分离性能相对较差。在欠定情况下, 混合矩阵的估计通常利用一种“直线聚类”的方法来完成, 直线的方向向量即为混合向量的一个估计, 此类方法中比较有代表性的是 K-PCA 方法^[5](即 K 均值聚类与主分量分析相结合的方法)。但是直线聚类方法依赖语音信号的近似 W-分离正交性, 当满足 W-分离正交的时频点很少时, 其估计性能严重下降。

针对上述讨论, 本文提出一种新的基于稀疏性的混合矩阵估计方法。该方法利用主分量分析(Principal Component Analysis, PCA)检测出只有一个源信号存在的时频点, 然后用于估计混合矩阵。由于检测出的时频点满足 W-分离正交, 因

此提高了估计精度, 克服了源信号增多导致近似 W-分离正交性变差所带来的影响。同时本文研究了语音信号的稀疏性对分离性能的影响, 指出了导致两类基于稀疏性的欠定语音盲分离方法失败的原因。最后通过仿真对上述结论进行了验证。

1 问题描述

N 个语音源信号和 M 个传感器的实时线性混合模型为:

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

其中: $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$ 和 $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$ 分别表示观测信号和语音源信号, $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N]$ 为 $M \times N$ 的混合矩阵, $\mathbf{a}_i (i = 1, \dots, N)$ 为混合向量, 其元素均为常数。盲源分离是指仅根据观测信号 $\mathbf{x}(t)$ 估计出源信号 $\mathbf{s}(t)$, 本文只考虑欠定混合情况 ($M < N$)。

2 语音信号的稀疏性

2.1 W-分离正交

语音信号在时频域呈现出较好的稀疏性, 因此通过 L 点的短时傅里叶变换(Short Time Fourier Transform, STFT) 转换到时频域表示:

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

设 $S_i(\omega, t)$ 和 $S_j(\omega, t)$ 是任意两个语音信号的时频表示, 则 W-分离正交^{[2]529} 描述为:

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

也即是说每个时频点上至多存在一个源信号。W-分离

收稿日期: 2008-10-23; 修回日期: 2008-12-15。基金项目: 国家自然科学基金资助项目(60672157; 60672158)。

作者简介: 王国鹏(1983-), 男, 河北沧州人, 硕士研究生, 主要研究方向: 语音信号的盲分离; 刘郁林(1971-), 男, 四川简阳人, 教授, 博士, 主要研究方向: 盲信号处理、超宽带通信、认知无线电、无线传感器网络、DSP; 罗颖光(1984-), 男, 河南驻马店人, 硕士研究生, 主要研究方向: 超宽带通信系统。

正交的假设使得问题简化,但实际上这种假设并非在所有时频点上均成立,所以称语音信号服从近似 W-分离正交性。

2.2 稀疏性的度量

在混合信号中,为了衡量源信号的近似 W-分离正交性,引入一个参数^{[2]531}:

$$r_k(z) = \frac{\sum_{(\omega,t)} |\Phi_{(k,z)}(\omega,t) X_{jk}(\omega,t)|^2}{\sum_{(\omega,t)} |X_{jk}(\omega,t)|^2} \times 100\% \quad (4)$$

其中, $X_{jk}(\omega,t) = STFT[a_{jk}s_k(t)]$ 表示参考传感器 J 上第 k 个源信号的短时傅里叶变换。 $\Phi_{(k,z)}(\omega,t)$ 是一个与参数 z 有关的时频掩蔽:

$$\Phi_{(k,z)}(\omega,t) = \begin{cases} 1, & 20 \log(|X_{jk}(\omega,t)| / |\hat{X}_{jk}(\omega,t)|) > z \\ 0, & \text{其他} \end{cases} \quad (5)$$

其中 $\hat{X}_{jk}(\omega,t) = STFT\left[\sum_{i=1, i \neq k}^N X_{ji}(t)\right]$ 表示参考传感器 J 上其他源信号的干扰。 $r_k(z)$ 越大,表明信号的稀疏性越强。

3 基于 PCA 的混合矩阵估计

随着混合信号数目的增加,语音信号的近似 W-分离正交性越来越差,因此基于稀疏性的混合矩阵估计方法的性能也随之下降,甚至有可能失败。本文提出了一种基于 PCA 的混合矩阵估计方法,通过检测只有一个源信号存在的时频点,以保证用于估计混合矩阵的时频点满足 W-分离正交,从而提高了估计精度。

3.1 混合信号的 PCA

假设观测信号 $X(\omega,t)$ 的均值为零,其协方差矩阵为:

$$\mathbf{R}_x(\omega,t) = E[X(\omega,t)X(\omega,t)^H] = A\Sigma A^H \quad (6)$$

其中 $\Sigma = \text{diag}[r_{11}, r_{22}, \dots, r_{MM}]$ 为一对角矩阵。对上述协方差矩阵进行特征值分解:

$$\mathbf{R}_x(\omega,t) = UAU^H \quad (7)$$

其中 $U = [\mathbf{u}_1, \dots, \mathbf{u}_M]$ 和 $A = \text{diag}[\sigma_1^2(\omega,t), \dots, \sigma_M^2(\omega,t)]$ 分别为特征向量矩阵与特征值矩阵, $\sigma_i^2(\omega,t)$ ($i = 1, 2, \dots, M$) 表示第 i 个源信号的方差(信号功率), \mathbf{u}_i 是其对应的特征向量, $(\cdot)^H$ 表示共轭转置。假设语音信号在该时频点处满足 W-分离正交,并且只有第 m 个源信号存在,而其他信号的功率均为零:

$$\sigma_m^2(\omega,t) \neq 0, \sigma_i^2(\omega,t) = 0; i = 1 \sim M, i \neq m \quad (8)$$

式(7)可以简化为:

$$\mathbf{R}_x(\omega,t) = \mathbf{u}_m \sigma_m^2(\omega,t) \mathbf{u}_m^H \quad (9)$$

如果只有第 m 个源信号存在,式(2)退化为:

$$X(\omega,t) = \mathbf{a}_m S_m(\omega,t) \quad (10)$$

其协方差矩阵:

$$\mathbf{R}_x(\omega,t) = \mathbf{a}_m E[S_m(\omega,t)S_m(\omega,t)^H] \mathbf{a}_m^H = \mathbf{a}_m \mathbf{r}_{mm} \mathbf{a}_m^H \quad (11)$$

比较式(9)和式(11)不难看出, \mathbf{u}_m 即是混合向量 \mathbf{a}_m 的一个估计,二者的幅度成比例。

3.2 满足 W-分离正交的时频点检测

由 3.1 节可知,对于满足 W-分离正交的时频点,最大特征值对应的特征向量即为一个混合向量的估计,如果考虑到噪声的影响,式(6)的特征值满足:

$$\sigma_m^2(\omega) \gg \sigma_i^2(\omega); i = 1, 2, \dots, M, i \neq m \quad (12)$$

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

$$\varepsilon = 1 - \sqrt{\sigma_2^2} / \sqrt{\sigma_1^2} \quad (13)$$

当 ε 很大时,说明其中某个源信号的能量非常大,此时只有一个源信号存在,说明该时频点满足 W-分离正交。

3.3 混合矩阵的估计

对于检测出的单一源信号时频点,其最大特征值对应的特征向量即为一个混合向量的估计。为了使这些特征向量中包含了对所有混合向量(N 个)的估计,本文做如下定义和假设:

定义 1 如果一个源信号在时频域中至少存在一个单一源信号时频点,则称这个源信号是“可见的”。

假设 在时频域中,所有源信号都是“可见的”。

通过上述假设,保证了每个混合向量至少被估计一次。因此,通过检测得到的特征向量是 N 个混合向量估计的集合,利用 K 均值聚类将该集合分为 N 类,聚类中心即为混合向量的最终估计,从而估计出混合矩阵。

3.4 提出方法的检验

取 $M = 2, N = 2 \sim 5$, 混合向量对应的方向角度分别为 $0^\circ, 30^\circ, 60^\circ, 90^\circ$ 和 135° , 语音源信号采用 IEEE 推荐信号 (<http://www.utdallas.edu/~loizou/speech/noizeus/clear.zip>), 时长 2.5 s, 采样频率为 8 kHz, 检测出的时频点的散点图如图 1 所示。其中, X_1 和 X_2 表示两个传感器上观测信号对应的坐标。可以看出,经检测后的时频点具有清晰的直线特点,因此能够提高混合矩阵的估计精度。

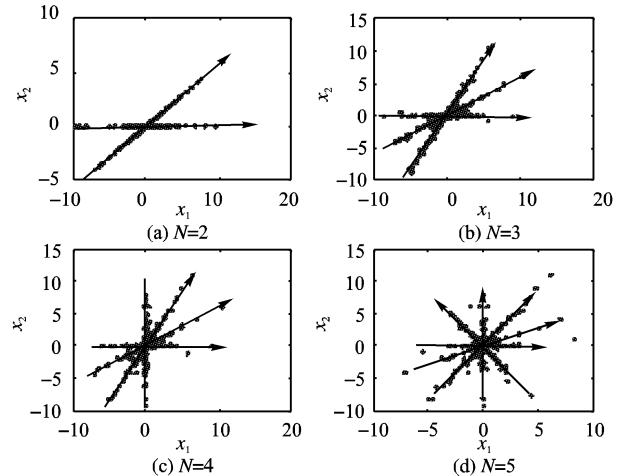


图 1 经检测后的散点

4 源信号的重构

在估计出混合矩阵以后,结合信号子空间方法^[3]重构源信号。信号子空间法假设每个时频点上存在的源信号数目不超过 $M - 1$, 在一定程度上降低了近似 W-分离正交性变差带来的影响,并且计算复杂度低于常用的最短路径法。最后对时频域信号 $y_k(\omega,t)$ 进行短时傅里叶逆变换 (Inverse Short Time Fourier Transform, ISTFT) 得到时域信号 $y_k(t)$ 。

5 实验仿真

本文采用三个传感器,分别对源信号数目 $N = 3, 4, 5, 6$ 的情况进行了仿真。语音源信号同 3.4 节,取 $L = 512$, STFT 采用 Hamming 窗,估计阈值 $\varepsilon = 0.96$ 。采用如下混合矩阵(当源信号的数目为 N 时,取矩阵 A 的前 N 列):

$$A = \begin{bmatrix} -0.8656 & -0.4519 & -0.0474 & -0.0861 & -0.5338 & -0.9234 \\ -0.0272 & -0.3415 & -0.1491 & 0.2707 & 0.4034 & 0.0290 \\ 0.5000 & 0.8241 & 0.9877 & 0.9588 & 0.7431 & 0.3827 \end{bmatrix} \quad (14)$$

定义两个性能评价指标。首先定义混合矩阵估计的性能评价指标,它表述为原矩阵和估计矩阵之间的广义干扰误差^[6](Generalized Crosstalk Error, GCE) :

$$GCE = \min_{M \in H} \|A - \hat{A}M\| \quad (15)$$

其中, H 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{A} 是 A 的估计,二者相等时 GCE 为零。

其次定义信号分离的性能评价指标,表示为信干比^[2](Signal to Interference Ratio, SIR) :

$$SIR_i = 10 \log \left(\frac{\sum_t y_i^s(t)^2}{\sum_t y_i^f(t)^2} \right) \quad (16)$$

其中, $y_i^s(t)$ 是分离信号 $y_i(t)$ 中来自于源信号 $s_i(t)$ 的成分, $y_i^f(t)$ 是来自于其他源信号的干扰成分。 SIR_i 越大分离效果越好。

首先比较本文方法与 K-PCA 方法的混合矩阵估计性能,仿真结果如表 1 所示。由于混合信号的近似 W-分离正交性较差,满足 W-分离正交的时频点很少,所以 K-PCA 方法的估计精度不高,并且在 $N = 5$ 和 $N = 6$ 时估计失败。由于本文提出的方法利用了满足 W-分离正交的部分时频点,所以估计性能得到了提高,即使在 $N = 5$ 和 $N = 6$ 时估计误差也较小,具有很好的鲁棒性。

表 1 混合矩阵估计性能/GCE

估计方法	$N = 3$	$N = 4$	$N = 5$	$N = 6$
K-PCA	0.9269	1.4004	/	/
本文方法	0.0214	0.0846	0.1516	0.1701

其次讨论源信号的近似 W-分离正交性对分离性能的影响。源信号的数目不同时,各源信号的近似 W-分离正交性如表 2 所示,本文方法的分离性能如表 3 所示。从表 2~3 中可以看出,随着混合信号数目的增加,源信号的近似 W-分离正交性越来越差,而分离性能也随之下降。在 $N = 6$ 时,第一个源信号和第六个源信号分离效果很差,因为其近似 W-分离正交性较差;即使采用原始混合矩阵 A ,其分离性能分别为 4.43 dB 和 11.58 dB,仍然没有达到较好的分离效果,实际试听效果差。因为随着混合信号数目的增多,各语音信号时频点之间的重叠逐渐增多,能量较小的语音信号的部分时频点被能量较大的语音信号掩蔽掉,导致其近似 W-分离正交性下降,当低于某个值时,即使能够估计出混合矩阵,也未必能分离出该源信号。

表 2 源信号的稀疏性 $r_k(z)$ %

源信号数 N	S_1	S_2	S_3	S_4	S_5	S_6
3	63.46	87.81	68.86	/	/	/
4	54.05	82.79	51.08	70.72	/	/
5	43.02	75.03	46.06	58.64	77.31	/
6	40.32	66.81	43.14	50.25	73.69	37.62

表 4 列出了时频二元掩蔽方法中比较有代表性的 BLUES 方法^[4]的分离性能。对比可以看出,本文方法的分离性能优于 BLUES 方法。因为 BLUES 方法假设每个时频点上至多存在一个源信号,通过提取属于同一个源信号的时频点来分离信号,而信号子空间方法在重构源信号时,允许每个时频点

上至多存在 $M - 1$ 个源信号,所以 BLUES 方法对近似 W-分离正交性的依赖性更强,当近似 W-分离正交性变差时,其分离性能下降更快。综上所述,基于稀疏性的语音盲分离方法与源信号的近似 W-分离正交性密切相关,但需要指出,本文的结论只是说明了一个事实,仿真结果不适用于所有情况,因为近似 W-分离正交性与源信号的稀疏性以及混合矩阵有关,可分离的源信号的数目也随之变化,而更复杂的情况有待进一步探讨。

表 3 本文方法的估计性能 SIR dB

源信号数 N	y_1	y_2	y_3	y_4	y_5	y_6
3	42.01	42.55	42.79	/	/	/
4	33.95	28.56	33.42	37.03	/	/
5	27.45	26.43	40.17	27.32	30.68	/
6	0.94	19.77	16.47	18.33	18.88	4.10

表 4 BLUES 方法的估计性能 SIR dB

源信号数 N	y_1	y_2	y_3	y_4	y_5	y_6
3	22.85	19.09	31.50	/	/	/
4	19.39	16.29	18.25	8.01	/	/
5	11.72	13.69	8.91	13.41	19.17	/
6	/	13.03	8.77	1.96	7.37	2.08

6 结语

为了能更好地利用语音信号的稀疏性,本文提出了一种基于主分量分析的混合矩阵估计方法,提出的方法利用满足 W-分离正交的时频点估计混合矩阵,不受源信号数目增多导致近似 W-分离正交性下降带来的影响,具有较好的鲁棒性。但是当近似 W-分离正交性下降到一定程度时,就无法估计源信号,因此基于稀疏性的欠定语音盲分离具有局限性。

参考文献:

- [1] MAKINO S, LEE T W, SAWADA H. Blind speech separation[M]. Berlin: Springer, 2007.
- [2] 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, D C: IEEE Computer Society, 2002, 1: 529~532.
- [3] AÏSSA-EL-BEY A, ABED-MERAIM K, GRENIER Y. Underdetermined blind source separation of audio sources in time-frequency domain[C]// Proceedings of the 2005 Signal Processing with Adaptive Sparse Structured Representations. Rennes, France: [s. n.], 2005: 115~118.
- [4] PEDERSEN M S, WANG D L, LARSEN J, et al. Overcomplete blind source separation by combining ICA and binary time-frequency masking[C]// Proceedings of the 2005 IEEE International Workshop on Machine Learning for Signal Processing. Washington, D C: IEEE Computer Society, 2005: 15~20.
- [5] 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.
- [6] THEIS F J, LANGA E W, PUNTONET C G. A geometric algorithm for overcomplete linear ICA[J]. Neurocomputing, 2004, 56(1/4): 381~398.