

改进的周期信号盲抽取算法^①

杜 瑞, 王剑书

(西北工业大学 电子信息学院, 西安 710072)

摘 要: 周期信号的盲抽取在医学等领域具有重要的研究价值. 对于此类信号的抽取需要利用先验知识对其周期进行较为准确的估计, 多数算法对于目标信号周期估计误差十分敏感. 文章针对此问题提出一种改进的周期特征信号盲抽取算法, 首先通过先验知识估算最佳时延的存在区间并利用改进的时域相关函数对提取矩阵进行初始估计, 再使用负熵作为目标函数并引入自相关函数确保抽取顺序, 使用牛顿迭代法进行优化, 达到最优结果. 仿真和实际 ECG 数据测试表明, 对比几种现行算法本文提出的算法抽取效果较好, 并增强了周期估计误差的鲁棒性.

关键词: 盲抽取; 时域相关函数; 负熵; 牛顿迭代法

Improved Algorithm for Extracting Periodic Signals

DU Rui, WANG Jian-Shu

(School of Electronic and Information, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: The research of extracting periodic signals is of great value in the field such as medicine. A precise estimation for period of target signals with priori knowledge is necessary in such kind of algorithms. But most of algorithms are sensitive to the error from the estimation. The algorithm proposed in the article including steps as follows: first roughly evaluate the extracting matrix with the improved time domain correlation function by evaluating the existing interval of optimizing time delay, use negentropy as the target function with the correlation function given into consideration to ensure the order and optimize it with Newton iterative method. The simulation and the experiment with real ECG data prove that the proposed algorithm can get a better result in the extraction than other algorithms, and show its robustness to the error of time delay.

Keywords: BSE; time domain correlation function; negentropy; Newton iterative method

盲抽取(Blind Source Extraction, BSE)是一种利用先验知识从混合信号中抽取一个或者一部分感兴趣信号的方法. 相比盲分离(Blind Source Separation, BSS)方法, BSE 在提取感兴趣信号时具有明显优势. BSE 利用感兴趣信号的先验知识, 直接抽取目标信号, 避免了 BSS 中信号幅度和次序不确定性的问题. 近年来, 周期特性信号的盲抽取问题成为 BSE 领域中的一个研究热点, 在心电图(Electrocardiogram, ECG)、脑电图(Electroencephalogram, EEG)和功能磁共振成像(Functional Magnetic Resonance Imaging, FMRI)等应用中具有较高的价值.

Barros A K 等^[1]在 2001 年提出了一种对感兴趣准周期信号的快速抽取算法, 但该算法高度依赖于周期估计的精度. Zibulevsky M 等^[2]在 2002 年提出一种利用源信号稀疏特性对目标信号进行分离的盲抽取算法, 但由于计算量过大, 限制了其应用. Zhang Zhilin 等在 2005 年提出一种利用延时相关矩阵直接获取分离向量的方法^[3], 在 2006 年提出一种两步法^[4], 但该算法仍然对周期估计的精度较为敏感. 同一时期, Lu Wei 等^{[5][6]}提出了参考 ICA 法, 使用拉格朗日函数优化策略, 对输出源的负熵进行优化, 完成了几路源信号的抽取, 但这种方法存在收敛不稳定性. Shi Zhenwei

① 收稿时间:2013-10-28;收到修改稿时间:2013-11-25

等^[7]在 2007 年通过设定所提取信号的某个可微函数的自相关函数作为目标函数, 实现了盲抽取; Li Xilin 等^[8]提出对各观测通道信号的二阶相关矩阵做最小化解相关处理的目标函数, 采用逼近联合对角化技术同样实现了盲抽取, 这两种算法虽然从不同方面解决了盲抽取问题, 但并没有避开对周期估计误差敏感这一问题. Zhang Hongjuan 等^[9]在 2009 年提出一种基于广义自相关和非高斯性的方法, 该方法实质为多目标优化问题的加权解法, 算法性能对权值较为敏感. Li Changli 等^[10]2010 年提出一种基于负熵的盲抽取算法, 使用信号自相关函数与负熵的积作为优化函数, 该方法需要目标信号的自相关存在并且达到峰值(对周期有最佳估计)时才能得到较好的抽取结果. 陈寿奇等^[11]2011 年提出一种改进的盲抽取算法, 对周期的估计误差具有一定的鲁棒性, 但是只针对最佳周期估计邻域内的一些离散点才具有较好的抽取效果.

以上算法均需要运用先验知识对周期做比较准确的估计. 针对此问题, 本文提出一种改进的周期信号抽取算法, 对周期的估计误差有较强的鲁棒性, 且具有更高的准确度.

1 盲抽取问题

1.1 盲抽取问题描述

BSE 并不要求分离所有混合的 n 路源信号, 只在采集到的 m 路混合信号中抽取感兴趣的 l 路源信号, ($l < m, n$). 其数学模型如下: 假设源信号有 n 路, 构成源信号矩阵, 即 $S = [s_1, s_2, \dots, s_n]^T$. 经混合矩阵 $A(m \times n)$ 进行混合, 我们假设 X 为混合后的矩阵, $X = A \times S = [x_1, x_2, \dots, x_m]^T$, x_1, x_2, \dots, x_m 为观测信号. 现在我们需要寻找一个抽取矩阵 $W(l \times m)$ 抽取 l 路信号, 使得 $Y = W \times X = W \times A \times S$. 则 Y 就是我们使用现有条件观测信号 X 和抽取矩阵 W 求得源信号 S 的估计值. 通过比较 Y 和 S 的近似程度, 在 Y 的迭代过程中不断修正 W , 最终使其收敛, 得到 S 的最优估计.

白化是盲信号处理(Blind Source Processing, BSP)中最常用的预处理方法, 即通过一定的线性变换 $\bar{X} = VX$, 使得 $E\{\bar{X}\bar{X}^T\} = I$, V 为白化矩阵, 其取值如下: 设混合信号 X 的相关矩阵为 R_x , 由相关矩阵性质可知 R_x 存在特征值分解: $R_x = Q\Sigma^2Q^T$, Σ^2 为对角阵, 其元素为矩阵 R_x 的特征值, 正交矩阵 Q 的列向量为这些特征正值对应的标准正交的特征向量, 有

$V = \Sigma^{-1}Q^T$. 为了方便起见, 本文后面的 X 都指白化后的观测信号, 我们假设 $m = n$, 并且只讨论一路信号的抽取(此时 W 为 $1 \times m$ 维, 即为一向量, 下文我们以 w 表示这个分离向量).

盲信号抽取整体思路如下:

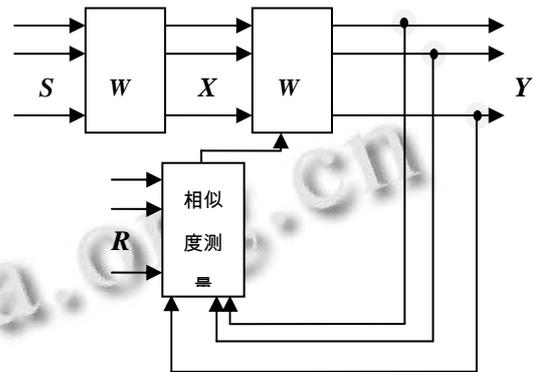


图 1 盲信号抽取整体思路

对混合信号进行抽取得到目标信号的估计, 结合先验知识 R 进行相似性度量, 判断收敛, 输出结果.

我们抽取的目标为周期信号, 那我们可以假设期望抽取的周期信号 s_i 对于特定的时延 τ 满足以下关系式:

$$\begin{cases} E\{s_i(k)s_i(k-\tau)\} > 0 \\ E\{s_j(k)s_j(k-\tau)\} = 0 (\forall j \neq i) \end{cases} \quad (1)$$

其中 k 为时间序列.

使用目标函数:

$$J(w) = E\{y(k)y(k-\tau)\} = w^T E\{X(k)X(k-\tau)^T\} w \quad (2)$$

其中 X 为混合信号, $E\{\cdot\}$ 表示对目标信号求期望.

在 $\|w\|=1$ 的限定条件下最大化目标函数 $J(w)$ 即可. 对于一个期望抽取的周期信号 s_i , 当 τ 为目标信号的周期时, $J(w)$ 达到最大值, 此时的 τ 称为最佳时延. 求得此时的 w , 便可以抽取目标信号.

1.2 盲抽取算法

两步法是一种对上述方法的改进, 也是现在比较常用的算法, 其先对抽取向量 w 进行粗估计, 观察目标方程:

$$J(w) = \frac{1}{2}J(w) + \frac{1}{2}J(w)^T = \frac{1}{2}w^T (R_x(\tau) + R_x(\tau)^T) w \quad (3)$$

其中, $R_x(\tau) = E\{X(k)X(k-\tau)^T\}$.

可以发现在限定条件 $\|w\|=1$ 下, 求解 w 与求解实对称矩阵 $(R_x(\tau) + R_x(\tau)^T)$ 的最大特征值对应的特征向

量是等效的. 第一步求出的 w^* 为 w 的一个粗糙估计, 将 w^* 作为初值带入第二步优化算法进行优化, 最后得到的就是 w 的最佳估计, 由 $y = w^T X$ 得到抽取信号. 此类算法对于 τ 的估计误差也具有很强的敏感性, 且第二步结合了 ICA 算法, 这就引入了抽取信号的不确定性.

2 改进周期特征信号盲抽取算法

2.1 求得 w^*

初值 w^* 对于 w 的迭代收敛是十分重要的, 若距离 w 过远, 则在优化过程中容易收敛到其他的局部极大值. 所以选取的 w^* 应该尽量靠近 w , 本文将之前的目标函数修改为以下形式:

$$J(w) = w^T \left\{ \sum_{p=1}^h \sum_{\tau=\tau_l}^{\tau_u} (R_x(p\tau) + R_x(p\tau)^T) \right\} w \quad (4)$$

则与其对应的求取初值 w^* 的算法为:

$$w = EIG \left(\sum_{p=1}^h \sum_{\tau=\tau_l}^{\tau_u} (R_x(p\tau) + R_x(p\tau)^T) \right) \quad (5)$$

其中 p 是一个正整数, 将 $\tau = [\tau_l, \tau_u]$ 设定为一个利用先验知识对于目标信号周期的一个粗估计.

改进的目标函数对目标信号的时域特性在多个周期内进行了叠加, 增强了对于估计误差的鲁棒性, 从而得到的 w^* 作为初值可以使抽取信号更好的收敛到目标信号.

2.2 优化 w^*

我们采用基于负熵的目标函数, 负熵的定义如下:

$$J(y) = H(y_{gauss}) - H(y) \quad (6)$$

其中 $H(y) = -\int f(y) \log f(y) dy$ 为信号 y 的熵. $H(y_{gauss})$ 为与 y 具有相同协方差矩阵的高斯随机向量的熵.

负熵的值总是非负的, 当且仅当 y 具有高斯分布, 负熵为零.

负熵的计算过于复杂, 我们一般使用负熵的近似计算, 最大化如下的目标函数, 以得到 w 的值.

$$J(w) \propto \left(E\{G(w^T X)\} - E\{G(v)\} \right)^2 \quad (7)$$

其中 $G(\bullet)$ 为非二次函数, v 为零均值单位方差的高斯变量, 本文中的 $G(\bullet)$ 选取 $G(u) = -\frac{1}{a_2} \exp(-au^2/2)$.

这种基于负熵的提取算法具有一定的缺陷, 此类算法不能确定输出信号的次序, 并不能确定提取的我们期望的信号. 所以还需进一步处理, 利用(4)式可以推出下式:

$$\frac{\partial J(w)}{\partial w} = 2w^T \overline{R_{xx}} \quad (8)$$

其中 $\overline{R_{xx}} = \sum_{p=1}^h \sum_{\tau=\tau_l}^{\tau_u} (R_x(p\tau) + R_x(p\tau)^T)$, 根据牛顿迭代法我们可以构造 w 的迭代公式:

$$w^+ = w + h \frac{\partial J(w)}{\partial w} = w + 2hw^T \overline{R_{xx}} \quad (9)$$

h 为步长因子. 上式对 w 进行迭代时引入了目标信号的时域自相关特性, 可以确保收敛到目标信号, 从而解决收敛不确定性.

2.3 改进的盲抽取算法步骤

(1)通过先验知识估计期望提取信号的周期使其落在 $\tau = [\tau_l, \tau_u]$ 的区间内.

(2)利用(5)式计算 w 的初始值 w^* .

(3)使用(7)和(9)式对 w^* 进行优化.

(4)判断约束条件 $\|w^+\| = 1$ 是否成立, 若成立, 则结束, 输出 $y = (w^+)^T X$, 若不成立, 返回步骤(3)继续优化, 直至成立.

3 仿真与实验

3.1 模拟数据

我们采用人工模拟三路周期信号加上两路高斯白噪声构成源信号. 三路信号分别为

$$\begin{cases} s_1 = \cos(2 \times \pi \times f_1 \times k + 1), f_1 = 0.042; \\ s_2 = \sin(2 \times \pi \times f_2 \times k + 6 \cos(2 \times \pi \times f_2 \times k)), f_2 = 0.002 \\ s_3 = \cos(2 \times \pi \times f_3 \times k + 2 \times \cos(\pi \times 0.042 \times k)), f_3 = 0.005 \end{cases}$$

其中 k 最大值为 3000.

则我们可得源信号:

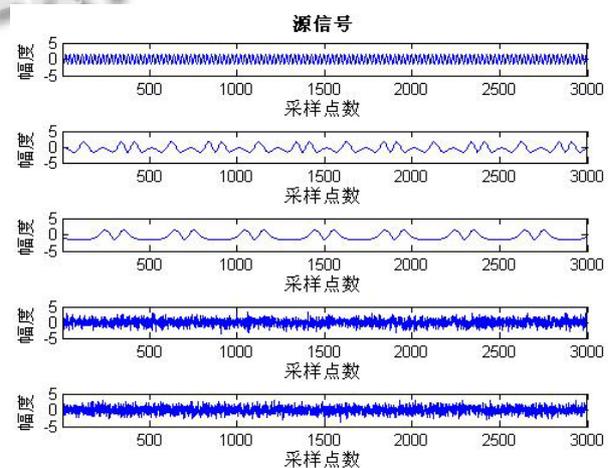


图 2 源信号

经过一个随机产生的 5×5 的混合矩阵可以得到混合后的信号:

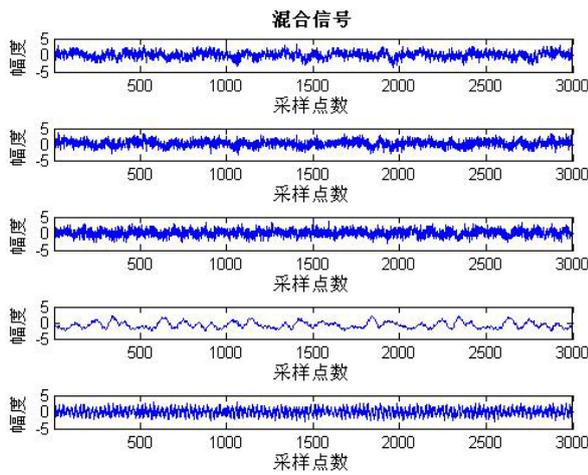


图 3 混合信号

现在我们期望抽取源信号 s_1 ，假设我们具备一定的先验知识，知道我们需要抽取的信号延时为 20 左右 (其准确值为 24)，则我们可以设定本文提出的算法中的 $\tau=[20,32]$ ，抽取 s_1 。为了进行对照特选取了其他四种算法进行比较，得到结果：

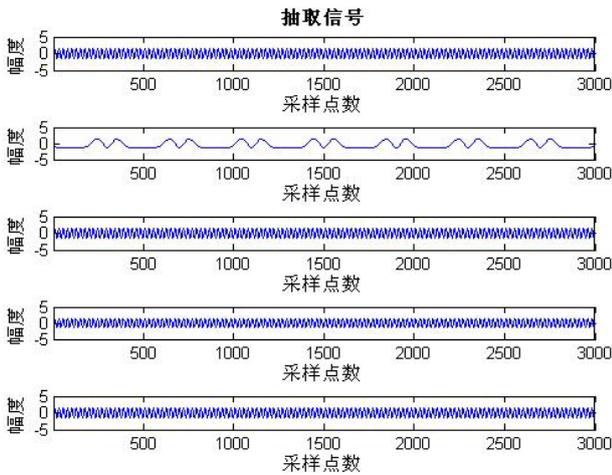


图 4 最佳时延抽取信号

自上向下使用的算法依次为文献[1]、[3]、[4]、[9]和本文提出的算法。图 4 中的 τ 取值依次为 $\tau=24$ ， $\tau=23$ ， $\tau=24$ ， $\tau=24$ ， $\tau=[20,32]$ 。图 5 中 τ 取值依次为 $\tau=30$ ， $\tau=21$ ， $\tau=31$ ， $\tau=29$ ， $\tau=[20,32]$ 。可见前几种算法对于 τ 非常敏感(图 4 中第二路的 τ 的估计只是略微偏差最佳时延，便产生错误的抽取结果)。

而且本文提取算法的准确率较高，使用性能指数

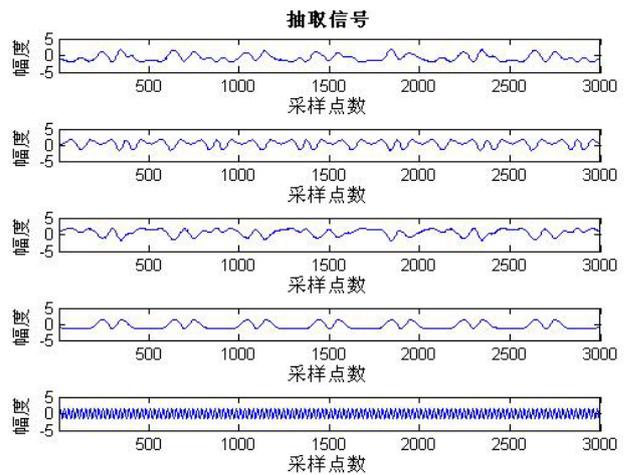


图 5 粗估计时延抽取信号

PI 作为度量(值越小性能越好)， $PI = \left(\sum_{j=1}^M \frac{|p_j|}{\max |p_k|} \right) - 1$ ，经过 50 次实验后求均值， \overline{PI} 值如表 1。

表 1 平均性能指标

算 法	[1]中	[3]中	[4]中	[9]中	本文
	算法	算法	算法	算法	算法
\overline{PI}	0.6072	0.3215	0.2431	0.0821	0.0255

3.2 ECG 数据实验

为了验证本算法的实际性能，我们采用 De Moor 在 1997 年给出的孕妇体表实际采集心电信号^[12]，此信号已经成为多种盲抽取的标准数据，用来检测各种方法的实际性能。

首先读入 ECG 信号进行显示，如下图：

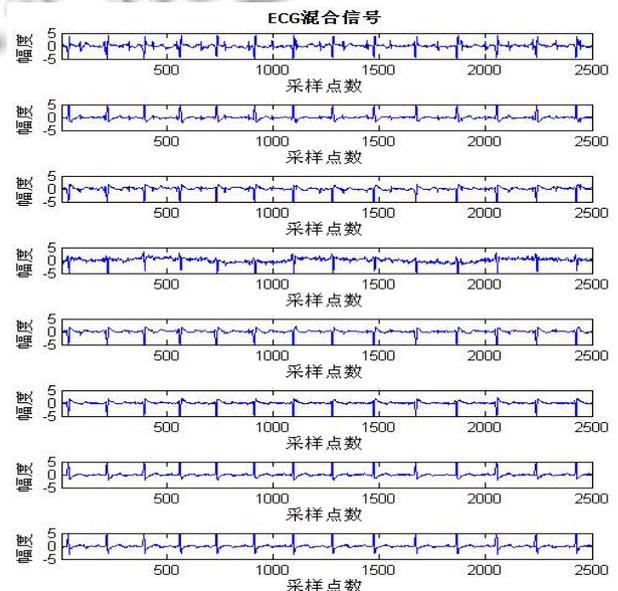


图 6 ECG 混合信号

选取多种算法与本文算法进行抽取并进行对比结果, 抽取结果如图:

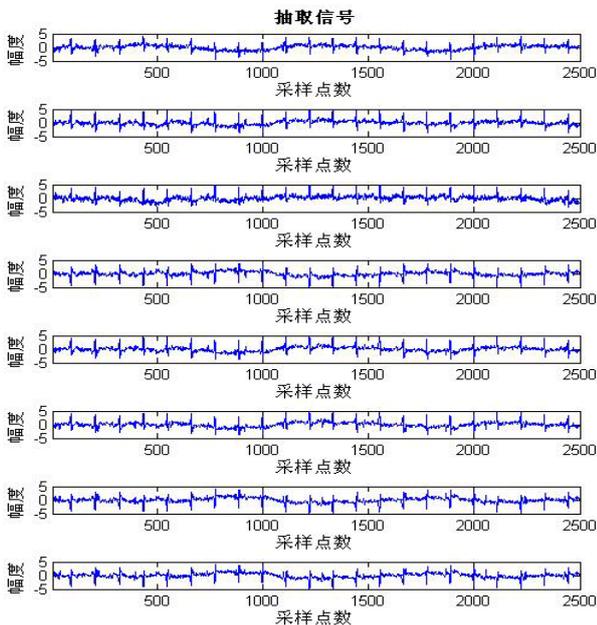


图 7 胎儿心电图抽取信号

图中自上而下第 1-4 路信号分别采用的算法为文献[1]、[3]、[4]、[9]中算法(都设定为最佳周期估计 $\tau=112$)得到的抽取信号, 第 5-8 路为本文提出的算法初步估计最佳时延, 设定 $\tau_1=[110,115]$, $\tau_2=[105,120]$, $\tau_3=[100,125]$, $\tau_4=[95,130]$ 分别得到的提取结果, 我们可以发现本文提出的算法较好的抽取出了期望的信号, 并且对 τ 的估计误差具有鲁棒性. 实际应用中, 我们可以直接观察图 4 中第一路信号中婴儿快而微弱的心跳信号, 可以初步估计目标信号的频率为 2 秒 4 次, 对应其周期为 120 左右, 设定如 τ_3 , τ_4 的估计区间即可得到较为理想的抽取结果.

提出的算法对在第一步计算 w^* 时采用了时域特性在区间内叠加求广义特征值得方法. 只要最佳时延落在区间当中, 最佳时延的信息就会存在于 w^* 之中, 经过迭代就可顺利得到抽取结果, 这就大大增强了估计误差的鲁棒性.

多次实验可以得到只要延时估计在落在大致为 [91,135] 的估计区间内(区间内的所有的估计子区间均可得到良好的抽取结果), 我们就可以得到良好的抽取结果.

改进的初始估计 w^* 能够更好的收敛到目标, 而牛

顿迭代法具有二阶收敛特性, 可以加快收敛速度.

经过多次计算后得到算法的抽取时间 T (其中[1]、[3]、[4]、[9]算法采用最佳时延估计)如表 2.

表 2 平均抽取时间

算 法	[1] 中 算 法	[3] 中 算 法	[4] 中 算 法	[9] 中 算 法	本 文 算 法
T	0.0091	0.0153	0.0232	0.0131	0.0104

采用不同的时延估计作为初值进行计算, 得倒的抽取时间如表 3.

表 3 本文算法不同时延估计的抽取时间

粗估计 时延	[110, 115]	[105, 120]	[100, 125]	[95, 130]
T	0.0075	0.0096	0.0102	0.0125

可以看出本文提出的算法, 在增强了信号延时估计误差鲁棒性的同时, 对计算速度也有了一定的提升.

4 结论

文章提出的算法从应用实际出发, 通过改进以时域相关函数为基础的粗估计方法, 较好的完成了对目标信号的盲抽取, 对需利用先验知识求得的最佳周期误差的鲁棒性有较显著的增强, 增加了算法的实际应用价值.

本文采用具有二阶收敛特性的迭代方法, 可以进一步优化加快收敛速度.

参考文献

- 1 Barros AK, Cichock A. Extraction of specific signals with temporal structure. *Neural Comput*, 2001, 13(9): 1995–2003.
- 2 Zibulevsky M, Zeevi YY. Extraction of a source from multi-channel data using sparse decomposition. *Neurocomputing*, 2002, 49 (1-4): 163–173.
- 3 Zhang Z, Zhang Y. Robust extraction of specific signals with temporal structure. *Neurocomputing*, 2005, 69: 888–893.
- 4 Zhang Z, Zhang Y. A two-stage based approach for extracting periodic Signals. *ICA, LNCS*, 2006, 3889: 303–310.
- 5 Lu W, Rajapakse JC. Approach and applications of constrained ICA, *IEEE Trans. on Neural Networks*, 2005, 16(1): 203–212.
- 6 Lu W, Rajapakse JC. ICA with reference. *Neurocomputing*,

- 2006, 69(16-18): 2244–2257.
- 7 Shi Z, Zhang C. Blind source extraction using generalized autocorrelations. *IEEE Trans. on Neural Networks*, 2007, 18(5) : 516– 1524.
- 8 Li X, Zhang X. Sequential blind extraction adopting second-order statistics. *Signal Processing Letters, IEEE*, 2007, 14(1): 58– 61.
- 9 Zhang H, Shi Z, Guo C. Blind source extraction based on generalized autocorrelations and complexity pursuit. *Neurocomputing*, 2009, 72: 2556–2562.
- 10 Li C, Guo S. A reference-based blind source extraction algorithm. *Neural Computing & Applications*, 2010, 19(2): 299–303 .
- 11 Chen S, Shen Y, Xu K. A two-stage blind extraction algorithm for fetal electrocardiogram. *Journal of Circuit and Systems*, 2011, 16(1): 125–130.
- 12 In: De Moor D, ed. *Daisy: Database for the identification of systems*. <http://www.esat.kuleuven.ac.be/sista/dais>.

www.c-s-a.org.cn

www.c-s-a.org.cn