

偏亚高斯语音信号有效分离ICA方法研究

洪 英¹, 韩周安²

(1. 北京理工大学信息科学技术学院 北京 海淀区 100081; 2. 电子科技大学电子工程学院 成都 610054)

【摘要】针对扩展Infomax语音分离算法仅只对无偏信号的概率密度分布进行建模的不足,提出了一种有效分离的有偏亚高斯信号ICA算法。通过修改扩展Informix算法所基于的Pearson混合模型,使修改后的模型既能较好地逼近对称的概率密度分布,又能逼近非对称的概率密度分布,从而在源信号是非对称分布的情况下,能获得更好的分离质量和较快的收敛速度。

关键词 语音分离; ICA算法; 独立分量分析; Infomax算法;
中图分类号 TN 912.16 文献标识码 A

Method Research on Effective Separated ICA Algorithm to Sub-Gaussian Distribution Audio Signal

HONG Ying¹ and HAN Zhou-an²

(1. College of Information Science and Technology, Beijing Institute of Technology Haidian Beijing 100081;

2. School of Electronic Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract An effective separated independent component analysis (ICA) algorithm is obtained by introducing a skewness-adjusting parameter to the Pearson mixture density model in extended Infomax algorithm. This model with skewness-adjusting parameter can cover a wider range of sub-Gaussian distribution including asymmetrical and multi-modal ones, resulting in more precisely approximating source's density. When dealing with non-skewed mixed sources, the new algorithm can achieve less steady-state error while maintaining fast convergence speed.

Key words audio separation; ICA algorithm; independent component analysis; infomax algorithm

独立分量分析(independent component analysis, ICA)是近十年发展起来的信号处理技术^[1],其基本研究思路是以非高斯信号为研究对象,在独立性假设的前提下对多路观测信号进行盲信号分离;在满足一定的条件下,能从多路观测信号中较好地分离出隐含的源信号。

噪声和干扰环境中的语音识别问题是困扰语音识别系统的难题,如果能实现噪声和语音的自动分离,在识别前就获得较为纯净的语音,可以彻底地解决噪声问题。

近年来取得很大进展的盲源分离为噪声和语音的分离提供了可能。在语音分离问题上,ICA算法是目前最好的一种技术,但它还不能进行实际的应用^[2]。用ICA进行语音分离比较困难^[3],因为语音分离可以建模成噪声混合的ICA模型、多通道卷积混合的ICA模型、非平稳信号混合的ICA模型、源信号个数动态变化的模型及混合矩阵动态变化的模型等。针对实际情况需要将上述的几个模型同时结合

起来进行研究。

针对当源信号既有超高斯信号又有亚高斯信号时,如何同时分离的问题,文献[4]提出了Infomax在线迭代学习算法,成功地对10个混合语音进行了分离。但是该算法需要对矩阵求逆,收敛速度慢,算法的性能要受源信号的混合程度的影响^[5];并且该算法只能分离超高斯信号。文献[6]在最大似然框架下提出了基于自然梯度的扩展Infomax算法,形式上更为简单,但在语音分离问题中源信号是完全未知的^[7]。该算法需要已知源信号的概率密度函数,必须找到一个合适的函数对源信号的概率密度函数进行逼近,即需要分别对亚高斯信号和超高斯信号的概率密度分布建模,从而得到2个非线性函数。但是在建模中,该算法没有考虑源信号分布不对称的情况,因此本文通过修改Pearson混合模型,提出了一种针对有偏亚高斯信号的有效分离ICA语音分离方法(ES-ICA),使得修改后的模型既能对对称分布的源信号建模,又能对分布不对称的源信号建模。

1 扩展Infomax算法

扩展Infomax算法通过在线切换不同的非线性函数来同时分离任意分布的源信号，是一个应用非常广泛的算法^[8]。

求解ICA问题的任务是估计ICA模型中的混合矩阵 \mathbf{A} 和 N 个独立的源信号 $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$ 。ICA模型为：

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

由随机向量分布密度的变换规则，并考虑到源信号之间的统计独立性，有：

$$p_x(\mathbf{x}) = |\det \mathbf{W}| \prod_{i=1}^N p_{s_i}(\mathbf{w}_i^T \mathbf{x}) \tag{2}$$

式中 $\mathbf{W} = \mathbf{A}^{-1}$ 为解混矩阵； $p_x(\mathbf{x})$ 为混合信号的联合概率密度； $p_s(\mathbf{s})$ 为源信号的联合概率密度； $p_{s_i}(s_i)$ 为源信号 s_i 的边缘概率密度。

设现有 T 次观测，记为 $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)$ ，则似然函数为：

$$L(\mathbf{W}) = \prod_{t=1}^T \prod_{i=1}^N p_{s_i}(\mathbf{w}_i^T \mathbf{x}(t)) |\det \mathbf{W}| \tag{3}$$

取对数并且两边同除以 T ，有：

$$\frac{1}{T} \ln L(\mathbf{W}) = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \ln p_{s_i}(\mathbf{w}_i^T \mathbf{x}(t)) + |\det \mathbf{W}| \tag{4}$$

从而得到了目标函数。对式(4)求关于 \mathbf{W} 的梯度，因此可得：

$$\frac{1}{T} \frac{\partial \ln L(\mathbf{W})}{\partial \mathbf{W}} = -E\{\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T\} + \mathbf{W}^{-T} \tag{5}$$

式中 $\mathbf{g}(\mathbf{u}) = [g_1(u_1), g_2(u_2), \dots, g_N(u_N)]^T$ ； $\mathbf{u} = [u_1, u_2, \dots, u_N]^T$ ； $g_i(u_i) = -\frac{p'_{s_i}(u_i)}{p_{s_i}(u_i)}$ 为非线性函数。

由此可得到基于一般梯度的学习算法为：

$$\Delta \mathbf{W} \propto -E\{\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T\} + \mathbf{W}^{-T} \tag{6}$$

式(6)的矩阵求逆会带来很大的计算量，并且导致收敛很慢。其收敛性能可以通过对观测信号进行白化以及使用自然梯度而得到改善。对式(6)的右端乘以 $\mathbf{W}^T \mathbf{W}$ ，去掉均值运算符，从而得到基于自然梯度的算法为：

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(\mathbf{I} - \mathbf{g}(\mathbf{u}(k))\mathbf{u}^T(k))\mathbf{W}(k) \tag{7}$$

至此，扩展Infomax算法的形式在最大似然估计的框架中推导出来。从式(7)可见，非线性函数与源信号的概率密度函数直接联系。如果已知源信号的概率密度函数 $p_{s_i}(i=1, 2, \dots, N)$ ，则可以得到非线性函数。

然而在ICA问题中不可能知道源信号的概率密度函数，但可以用某些模型来逼近源信号的概率分布^[9]，即可以选择一些合适的非线性函数来逼近真实的非线性函数。

所有的信号根据其峭度值可以分为高斯信号、亚高斯信号和超高斯信号三类。在ICA问题中，假设所分离的源信号中至多只有1个为高斯信号，只需考虑如何逼近亚高斯信号和超高斯信号的概率密度函数，可以采用Pearson混合模型(pearson mixture model)的对称形式来逼近服从亚高斯分布的概率密度函数。Pearson混合模型的对称形式为：

$$p(u) = \frac{1}{2}(N(\mu, \sigma^2) + N(-\mu, \sigma^2)) \tag{8}$$

式中 $N(\mu, \sigma^2)$ 为均值是 μ 、方差是 σ^2 的高斯分布函数。该函数在 $\sigma^2 = 1$ 、 $\mu = [0, 2]$ 变化的曲线如图1所示。当 $\mu = 0$ 时， $p(u)$ 为高斯分布；当 $\mu = 1.5$ 时， $p(u)$ 是一个双峰模型。

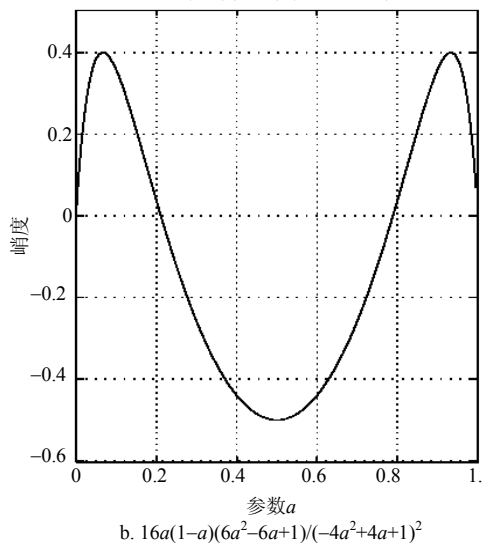
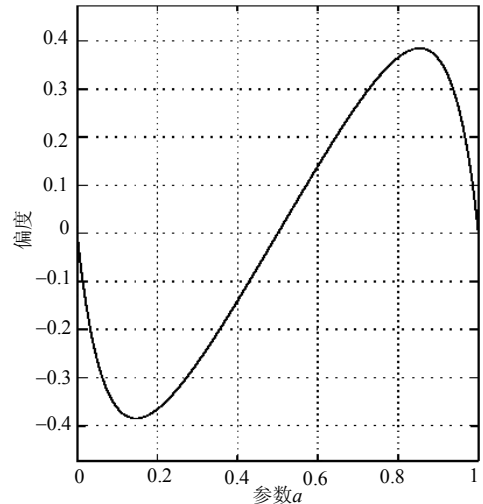


图1 非对称模型的偏度和峭度随 a 变化的情况

为了从理论上说明 $p(u)$ 的确可以用来逼近亚高斯信号的概率密度分布, 需要计算出它的归零化峭度。文献[10]给出了以下的计算结果:

$$k = \frac{c_4}{c_2^2} = \frac{-2\mu^4}{(\mu^2 + \sigma^2)^2} \quad (9)$$

式中 c_i 为 $p(u)$ 的第 i 阶累积量。当 $\mu = 0$ 时, $k = 0$ 。随着 μ 和 σ^2 的改变, 归零化峭度在 $-2 \sim 0$ 变换, 因而式(9)是一个亚高斯概率密度函数, 可以用来逼近任何亚高斯信号的真实分布。

2 ESICA算法

在语音分离中接触到很多信号都是分布不对称的有偏信号, 这些信号很多又属于亚高斯信号。如前所述, 扩展Infomax算法假设源信号的概率分布是对称分布的, 因此当分离的源信号为非对称分布时, 所采用的模型不能很好地逼近源信号真实的分布, 分离的效果将变差。

修改对称形式的Pearson混合模型式(8)为非对称形式, 可得:

$$p(u) = (1-a)N(\mu, \sigma^2) + aN(-\mu, \sigma^2) \quad (10)$$

则非线性函数为:

$$g(u) = -\frac{p'(u)}{p(u)} = \frac{u}{\sigma^2} - \frac{\mu}{\sigma^2} \left(\frac{1 + \frac{a}{a-1} e^{-\frac{2\mu u}{\sigma^2}}}{1 - \frac{a}{a-1} e^{-\frac{2\mu u}{\sigma^2}}} \right) \quad (11)$$

参数 a 的引入使该模型的偏度不再为零。随着 a 的取值范围从 $0 \sim 1$ 变化, 模型的偏度也不断变化。显然, 当 $a = 0.5$ 时, 它成为式(8), 偏度为零。因此, 如果要用式(8)逼近分布不对称的信号分布, a 的取值不能为零。

要确定 a 的最佳取值, 需要分析模型 $p(u)$ 的峭度和偏度。首先计算 $p(u)$ 的矩生成函数为:

$$\Phi(\varpi) = \int p(u)e^{j\varpi u} du = (1-a)e^{j\mu\varpi - \frac{\sigma^2\varpi^2}{2}} + ae^{-j\mu\varpi - \frac{\sigma^2\varpi^2}{2}} \quad (12)$$

从而得到累积量生成函数为:

$$\Psi(\varpi) = \ln(\Phi(\varpi)) = \ln \left[(1-a)e^{j\mu\varpi - \frac{\sigma^2\varpi^2}{2}} + ae^{-j\mu\varpi - \frac{\sigma^2\varpi^2}{2}} \right] \quad (13)$$

则 n 阶累积量为:

$$k_n = (-j)^n \left. \frac{d^n \Psi(\varpi)}{d\varpi^n} \right|_{\varpi=0} \quad (14)$$

归零化峭度为:

$$\tilde{k} = \frac{k_4}{k_2^2} = \frac{16a\mu^4(1-a)(6a^2 - 6a + 1)}{(4a\mu^2(1-a) + \sigma^2)^2} \quad (15)$$

当 $0.2113 < a < 0.7887$ 时, $\tilde{k} < 0$, 式(10)可用来逼近亚高斯分布。计算式(10)的偏度为:

$$S = \frac{k_3}{(k_2)^{3/2}} = \frac{8a\mu^3(1-a)(2a-1)}{(4a\mu^2(1-a) + \sigma^2)^{3/2}} \quad (16)$$

当 $0 < a < 0.5$ 时, $S < 0$, 式(10)可用来逼近偏度小于零的信号; 当 $0.5 < a < 1$ 时, $S > 0$, 式(10)可用来逼近偏度大于零的信号。模型的偏度和峭度随 a 变化的情况如图1所示。

为了对偏度小于零的亚高斯信号的分布建模, a 必须满足 $0.2113 < a < 0.5000$; 为了对偏度大于零的亚高斯信号的分布建模, a 必须满足 $0.5000 < a < 0.7887$ 。因此取 $a = 0.3750$ 来对偏度小于零的亚高斯分布建模, $a = 0.6250$ 对偏度大于零的亚高斯分布建模, 由此获得对扩展Infomax算法的修改, 即ESICA算法的非线性函数如下:

(1) 当偏度小于零时, 有:

$$g(u) = -\frac{p'(u)}{p(u)} = u - \frac{1 - 0.6e^{-2u}}{1 + 0.6e^{-2u}} \quad (17)$$

(2) 当偏度大于零时, 有:

$$g(u) = -\frac{p'(u)}{p(u)} = u - \frac{1 - 1.67e^{-2u}}{1 + 1.67e^{-2u}} \quad (18)$$

(3) 当偏度等于零时, 有:

$$g(u) = -\frac{p'(u)}{p(u)} = u - \tanh(u) \quad (19)$$

这是一种更适合分离有偏亚高斯信号的ICA算法, 比原来的扩展Infomax算法的适用范围更广。当源信号为对称(或近似对称)分布时, 它退化为一般的扩展Infomax算法; 当源信号为非对称分布时, 它根据偏度的正负采用不同的非线性函数, 以达到更小的稳态误差、更高的分离精度。该算法的描述如下^[11]: (1) 初始化分离矩阵 W , 通常取单位阵, 给步长 μ 设值。(2) 计算 $u(t) = Wx(t)$, 其中, $x(t) = [x_1(t), x_2(t), \dots, x_N(t)]^T$ 为 t 时刻的观测信号; $u(t) = [u_1(t), u_2(t), \dots, u_N(t)]^T$ 为 t 时刻的输出。(3) 计算 $k_i = E\{\sec^2 h^2(u_i(t))\} E\{u_i^2(t)\} - E\{\tanh(u_i(t))\} u_i(t)$ 和 $u_i(t)$ 的偏度 S_i 。当 $k_i > 0$ 时, 选择非线性函数 $g(u_i) = u + \tanh(u)$; 当 $k_i < 0$, $S_i > \eta$, η 为一很小的正数, 选择非线性函数式(18); 当 $k_i < 0$, $S_i < -\eta$, 选择非线性函数; 当 $k_i < 0$, $-\eta < S_i < \eta$, 选择非线性函数。(4) 分离矩阵 W 的迭代更新: $W(t+1) =$

$W(t) + \mu(I - g(u(t))u^T(t))W(t)$ 。(5) 令 $t \leftarrow t+1$ ，继续下一时刻的迭代。

3 仿真试验

本文对扩展Infomax算法和ESICA算法进行了比较，如图2~图4所示。在实验中增加了数据分离的难度，所用数据为语音数据，来自荷兰Eindhoven科技大学的语音数据库。在1个房间内的3个话筒录制，房间内有2男1女在说话，话筒录制声音为他们的自然混合声，各人独立说话的声音并未专门录制。由于3人的语音信号的独立性不如自然声音的独立性，因此分离难度加大。

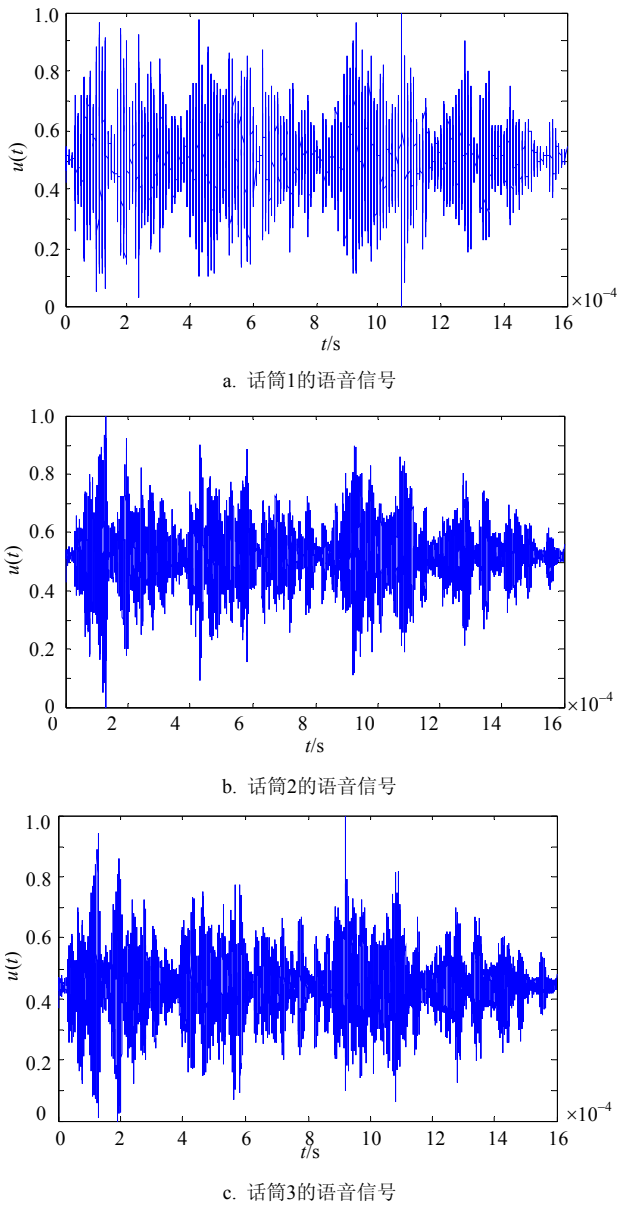


图2 各个话筒录制的混合语音信号

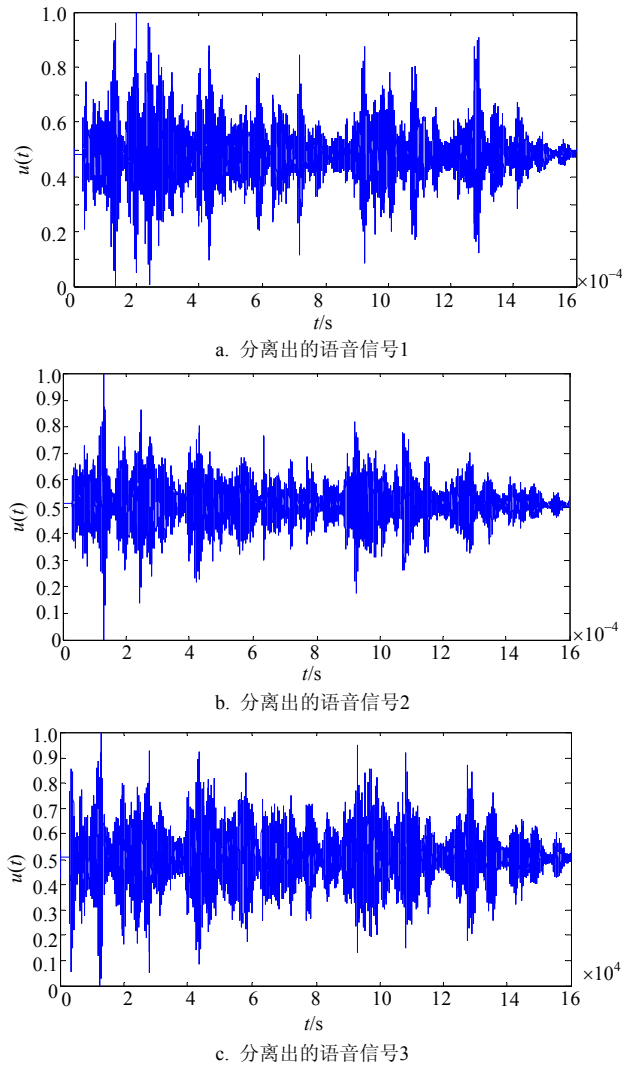
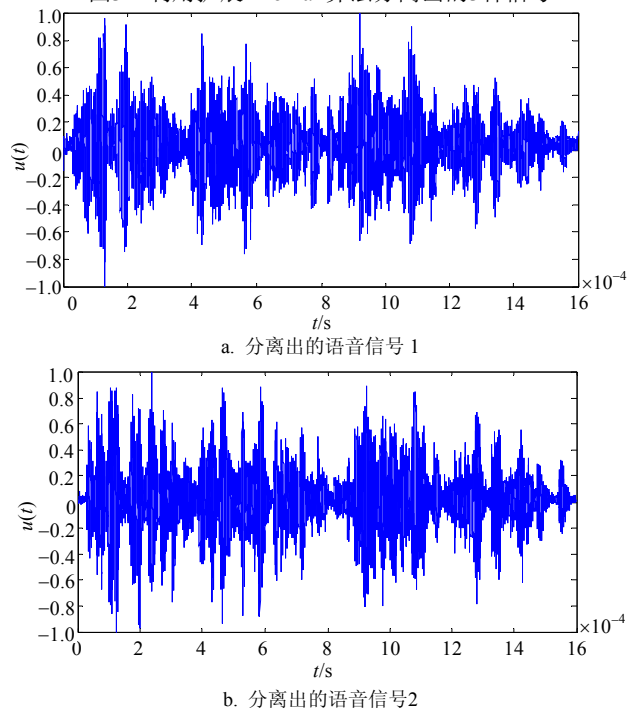
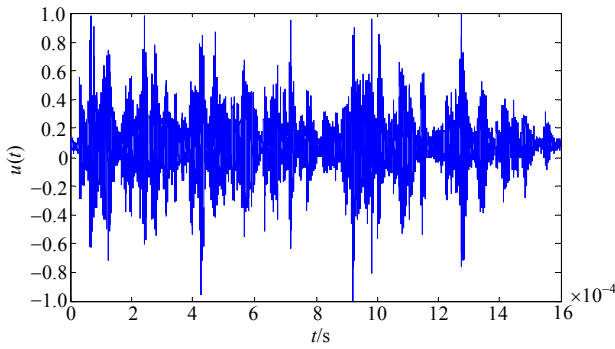


图3 利用扩展Infomax算法分离出的3种信号





c. 分离出的语音信号3

图4 利用ESICA算法分离出的3种信号

采用ESICA算法和扩展Infomax算法对图3所示的混合信号进行分离, 步长均取0.01。试验分别独立进行50次, 并取平均, 得到50次平均的性能指数, 如图4所示。在用2种算法分离之前, 首先做了去均值和白化的预处理。语音分离前后的峭度比较如表1所示。由表1可知, ESICA算法相比扩展Infomax算法具有更好的分离性能, 并且收敛速度损失不大。

表2 2种语音分离算法的性能比较

峭度	峭度		收敛步数		播放效果			
	Informax	ESICA	Informax	ESICA	Informax	ESICA		
分离前	语音1	5.140 2	分离后	语音1	5.814 9	7.268 2	分离效果稍差, 语音3中男女声混杂严重	语音能完全分离, 并且比较清楚
	语音2	5.679 8		语音2	8.012 1	10.432 1	48	62
	语音3	5.782 1		语音3	15.111 3	16.603 6		

4 结束语

EICA算法通过修改扩展Infomax算法所基于的Pearson混合模型, 使修改后的模型既能较好地逼近对称的概率密度分布, 又能逼近非对称的概率密度分布, 从而使在源信号是非对称分布的情况下, 能获得更好的分离质量和较好的收敛速度。而当源信号为对称分布时, 它又可以退化为一般的扩展Infomax算法。

参考文献

[1] TANG C W, VANSLYKE S A, CHEN C H. Electroluminescence of doped organic thin films[J]. J Appl Phys, 1987, 51: 913-915.

[2] HYVARINEN A, KARHUNEN J, OJA E. Independent component analysis[M]. New York: John Wiley & Sons, LTD, 2001.

[3] 官金安, 陈亚光. 基于独立分量分析的VEP中N2成分提取[J]. 计算机工程, 2006, 32(12): 29-31.

GUAN Jin-an, CHEN Ya-guang. Extraction of N2 components in visual evoked potentials based on independent component analysis[J]. Computer Engineering, 2006, 32(12): 29-31.

[4] LEE T W. Independent component analysis—theory and applications[M]. Norwell, MA: Kluwer, 1998.

[5] SAWADA Hiroshi, et al. Blind extraction of dominant target sources using ICA and time-frequency masking[J]. IEEE Trans on Audio, Speech and Language Processing, 2006, 14(6): 2165-2173.

[6] 张玲华, 杨震, 郑宝玉. 基于模糊分类器及多层前馈神经网络混合结构的说话人辨认[J]. 通信学报, 2005, (11): 35-37.

ZHANG Ling-hua, YANG Zhen, ZHENG Bao-yu. Hybrid architecture based on fuzzy classifier and multiplayer feed-forward neural network for speaker identification[J]. Journal on Communications, 2005, (11): 35-37.

[7] BELL A J, SEJNOWSKI T J. An information-maximization approach to blind separation and blind deconvolution[J]. Neural Computation, 1995, 7(6): 1004-1034.

[8] LEE T W, GIROLAMI M, SEJNOWSKI T J. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources[J]. Neural Computation, 1999, 11(3): 417-441.

[9] OZEROV A, et al. Adaptation of Bayesian models for single channel source separation and its application to voice/music separation in popular songs[J]. IEEE Trans on Audio, Speech and Language Processing, 2007, 15(5): 1564-1578.

[10] 郑鹏, 何同林, 刘郁林, 等. 基于实数编码遗传算法的盲信源分离方法[J]. 电子科技大学学报, 2006, 35(3): 295-297.

ZENG peng, HE Tong-lin, LIU Yu-lin, et al. Blind source separation method based on real coded genetic algorithm[J]. Journal of University of Electronic Science and Technology of China, 2006, 35(3): 295-297.

[11] 何致远, 胡起秀, 徐光祐. 两级决策的开集说话人辨认方法[J]. 清华大学学报(自然科学版), 2003, 43(04): 75-77.

HE Zhi-yuan, HU Qi-xiu, XU Guang-you. Method of open-set speaker identification with two-level decision strategy[J]. Journal of Tsinghua University(Science and Technology Edition), 2003, 43(04): 75-77.

编辑 黄莘