

基于脑电样本熵的测谎分析

高军峰^{1,2}, 司慧芳¹, 余彬¹, 顾凌云¹, 梁莹¹, 杨勇³

(1. 中南民族大学生物医学工程学院 & 认知科学国家民委重点实验室 & 中南民族大学医学信息分析及
肿瘤诊疗湖北省重点实验室 湖北武汉 430074;
2. 电子科技大学生命科学与技术学院 四川成都 610054;
3. 江西财经大学信息管理学院 江西南昌 330000)

摘要: 测谎分析在刑侦侦查和法律审判中具有重要意义. 为了区分是否说谎, 根据脑电信号的非线性特征, 本文首次使用非线性动力学的样本熵方法分析 30 名受试者处于诚实和说谎两种状态时脑电信号的样本熵值. 研究发现: 受试者处于诚实状态时的熵值波动范围明显小于说谎状态下的波动范围, 更重要的是说谎时的熵值显著高于说实话时的熵值, 表明样本熵可以区分诚实和说谎两种不同状态下的脑电信号, 该研究为基于脑电的测谎提供了一种新的途径.

关键词: 测谎; 脑电信号; 非线性特征; 样本熵

中图分类号: R338 **文献标识码:** A **文章编号:** 0372-2112 (2017)08-1836-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.08.005

Lie Detection Analysis Based on the Sample Entropy of EEG

GAO Jun-feng^{1,2}, SI Hui-fang¹, YU Bin¹, GU Ling-yun¹, LIANG Ying¹, YANG Yong³

(1. School of Biomedical Engineering, South-Central University for Nationalities & Key Laboratory of Cognitive Science, State Ethnic Affairs Commission & Hubei Key Laboratory of Medical Information Analysis and Tumor Diagnosis & Treatment, Wuhan, Hubei 430074, China;
2. School of Life Science and Technology, Chengdu, Sichuan 610054, China;
3. School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330000, China)

Abstract: There is great significance in lie detection for the criminal investigations and law trials. In this study, according to the nonlinear characteristics of electroencephalography (EEG), it is the first time to use the sample entropy (SE), a nonlinear dynamical parameter of EEG, to see if someone is lying. The sample entropy values of 30 subjects' EEG signals in lying or honesty states were calculated and analyzed. The study found that the fluctuating range of SE values in honesty was obviously less than that in lying. It is more important that the SE values in lying was significantly higher than the honesty, which indicated that SE could be used to distinguish EEG signals between two different states of honesty and lying. This research provides a new way for EEG-based lie detection.

Key words: lie detection; electroencephalography (EEG) signals; nonlinear characteristic; sample entropy

1 引言

撒谎是一种普遍存在的社会现象, 测谎在刑侦侦查和法律审判中具有重要意义. 传统的多道图测谎方法已经备受质疑, 近年来, 基于中枢神经系统活动的测谎方法迅速发展, 如: 功能性磁共振成像 (functional Magnetic Resonance Imaging, fMRI), 事件相关电位 (Event Related Potential, ERP) 等^[1,2]. 相对于 fMRI, ERP 具有较高的时间分辨率, 应用更为广泛. 目前认可的 ERP 产生机制有两种模型: 事件相关刺激诱发 EEG 相位重

排模型和刺激诱发新的叠加电位模型^[3]. ERP 中的 P300 成分与记忆和思维有关, 主要反映大脑认知加工过程, Rosenfeld 在利用 P300 进行测谎方面做了开创性的研究, 证明了用 P300 进行测谎的可行性^[4].

在基于 P300 的测谎方法中, 最重要的是提取 P300 成分及其特征. 当前对 P300 信号及其特征的提取方法主要有基于时域或频域的特征提取和相干平均^[5-7], 基于时域的特征提取方法忽略了脑电信号的频域特性, 特征较为简单, 基于频域的傅里叶分析方法利用信号频域的能量信息, 由于傅里叶变换的特点 (时域分辨率

最差),导致频率成分的时间定位信息损失^[8,9].自适应回归模型法对伪迹很敏感^[9],不适合分析非平稳随机信号.传统的相干平均法需多次刺激叠加,实验记录时间长,易引起受试者神经疲劳,不适合测谎研究^[10].目前使用较多的测谎方法有自举幅度差分(Bootstrapped Amplitude Difference, BAD)^[11],自举相关差分(Bootstrapped Correlation Difference, BCD)^[5]和模式识别三种方法,但是它们都存在一个共同的缺点:实际测谎的对象往往是嫌疑犯,犯罪场景和犯罪信息通常很少,而在仿真实验时,受试者需要接受大量刺激,容易引起受试者对少量信息的反复刺激采用反测谎策略.因此,采用单次或少次刺激的实验模式是很有必要的.

传统的时/频域的分析方法均基于对脑电产生系统的线性假设,即假设大脑是一个线性动力学系统,但脑电信号是大量神经细胞的非线性耦合,是一个高度非线性多单元连接的集合体,表现出明显的非线性特征^[12].随着非线性理论的进一步发展,基于非线性动力学系统的复杂度研究成为获取脑电特征的一种新兴方法.熵作为衡量时间序列复杂度的非线性动力学参数,已经在众多领域得到较好的应用. Richman 发现了一种非线性动力学分析方法——样本熵,该算法通过衡量时间序列中产生新模式概率的大小,得出脑电信号的复杂度,描述脑电信号的非线性特征.虽然基于样本熵的特征提取方法已经被用于分析脑电信号,但是在测谎领域还未见报道,本文尝试使用样本熵进行测谎分析,探究该算法在测谎研究领域的可行性.

2 样本熵理论

样本熵^[13]是一种非线性分析法,具有较好的一致性,通过度量信号的复杂度反映它的非线性特征,样本熵值越低,序列自我相似性越高,序列越简单,反之亦然.样本熵不但具备抗噪、抗干扰优点,而且避免了计算过程中因对自身数据量进行比较而导致的统计量不一致问题^[14].样本熵这种统计量的精确性使其适用于分析非线性的脑电信号和其他生物时间序列.具体算法步骤如下:

(1) 对于一个 N 点的时间序列,我们表示为:

$$\{\mathbf{u}(j) | 1 \leq j \leq N\} \quad (1)$$

(2) 构造一组 m 维向量:

$$\mathbf{X}_m(1), \mathbf{X}_m(2), \dots, \mathbf{X}_m(N-m) \quad (2)$$

其中,

$$\mathbf{X}_m(i) = \{\mathbf{u}(i+k) | 0 \leq k \leq m-1\} \quad (3)$$

(3) 定义任意两个 m 维向量之间的距离为:

$$d[\mathbf{X}_m(i), \mathbf{X}_m(j)] = \max\{|\mathbf{u}(i+k) - \mathbf{u}(j+k)| | 0 \leq k \leq m-1, i, j = 1 \sim N-m, j \neq i\} \quad (4)$$

(4) r 是设定的匹配过程的公差阈值,给定一个 r

值,然后对每个 i 统计 $d[\mathbf{X}_m(i), \mathbf{X}_m(j)]$ 中小于 $r \times SD$ (SD 为序列的标准差)数目 $N^m(i)$,并计算与距离总数的比值记作:

$$C_i^m(r) = N^m(i)/(N-m) \quad (5)$$

该过程称作 $\mathbf{X}_m(i)$ 模板匹配过程, $C_i^m(r)$ 表示任一 $\mathbf{X}_m(j)$ 与模板匹配的概率.

(5) $C_i^m(r)$ 的平均值为:

$$C^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m} C_i^m(r) \quad (6)$$

(6) 增加维数为 $m+1$, 构造一组 $m+1$ 维向量:

$$\mathbf{X}_{m+1}(1), \mathbf{X}_{m+1}(2), \dots, \mathbf{X}_{m+1}(N-m) \quad (7)$$

其中,

$$\mathbf{X}_{m+1}(i) = \{\mathbf{u}(i+k) | 0 \leq k \leq m\} \quad (8)$$

定义任意两个 $m+1$ 维向量之间的距离:

$$d[\mathbf{X}_{m+1}(i), \mathbf{X}_{m+1}(j)] = \max\{|\mathbf{u}(i+k) - \mathbf{u}(j+k)| | 0 \leq k \leq m, i, j = 1 \sim N-m, j \neq i\} \quad (9)$$

给定 r , 然后对每个 i 统计 $d[\mathbf{X}_{m+1}(i), \mathbf{X}_{m+1}(j)]$ 小于 $r \times SD$ 的数目 $N^{m+1}(i)$, 计算与距离总数的比值:

$$C_i^{m+1}(r) = N^{m+1}(i)/(N-m-1) \quad (10)$$

(7) 计算 $C_i^{m+1}(r)$ 的平均值:

$$C^{m+1}(r) = (N-m)^{-1} \sum_{i=1}^{N-m} C_i^{m+1}(r) \quad (11)$$

(8) 样本熵定义为:

$$SampEn(m, r) = \lim_{N \rightarrow \infty} \{-\ln[C^{m+1}(r)/C^m(r)]\} \quad (12)$$

当为有限值时,按上述步骤得出序列长度为 N 时

样本熵估计值. 记作

$$SampEn(m, r, N) = -\ln[C^{m+1}(r)/C^m(r)] \quad (13)$$

样本熵表示非线性动力学系统产生新信息的速率,样本熵值越低,产生新模式的速率越低.由以上定义可知,样本熵的值与模板长度 m 和公差阈值 r 有关,根据先验经验^[13],通常取 $m=2$ 或 3 , $r=0.1 \sim 0.25$.

3 实验过程与数据处理

3.1 测谎协议

实验挑选 30 名身体健康的本科生作为实验对象,平均年龄 21 岁,随机分为无辜和说谎两组,其中无辜者和说谎者各 15 名(两组男女比例均为 4:1).采用标准的三刺激模式测谎协议,实验前准备 6 个物品(仿真珠宝)及其对应的图片.针对说谎者,在事先准备好的保险箱中放入任意两个物品,要求受试者看清两个物品后拿走其中的一个,该物品的图片作为探针刺激(Probe, P),另外一个作为靶刺激(Target, T),其余物品作为无关刺激(Irrelevant, I);针对无辜者,保险箱中放入任意一个物品,要求受试者看清物品即可,该物品作为 T 刺激,任选其他一个物品作为 P 刺激,其余物品作为 I 刺激.脑电采集过程中,屏幕上随机显示 6 张物品

图片,要求受试者对看到的每一张图片做出反应(按鼠标键).要求说谎组看见 P 刺激时做出欺骗行为,即回答没有看过(按左键),出现其余两类刺激的时候要求说真话(T 刺激按右键,I 刺激按左键),无辜组全说真话^[15,16].

3.2 数据预处理

受试者脑部电极安放位置如下:F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz, O2,垂直眼电电极放在左眼上下 2 cm 处,水平眼电电极放在外眼角处,前额接地,右耳电极为参考电极.采用 Neuroscan 公司的 Synamps 放大器,通带频率为 0.3 ~ 30 Hz,采样率 500 Hz,各导联阻抗均小于 3k Ω ;E-Prime2.0 作为刺激软件.实验时 6 张图片将以 1600ms 的间隔随机出现 50 次,每张图片持续 500ms.每次实验中,I、P 和 T 刺激出现频率依次为 66.7%、16.7%和 16.7%.休息 5 min 后,再次重复上述实

验 2 次(共 3 次)^[6,15].

3.3 脑电去噪

已有学者证明,在 Pz 电极附近的 P300 非常明显^[6,16],且对两类受试者具有显著的区别意义,因此本文将 Pz 电极上的 P300 波形作为分析对象.使用 EEGLAB 工具箱^[8]对连续的 EEG 波形进行分割,将 P 刺激后 1300ms 分割为一个 Epoch.采用相干平均法去除噪声信号,将每名受试者在 Pz 电极上的每 6 个 P 刺激响应进行平均,得到两类受试者的 ERP 波形.最后两类受试者分别得到 375 ($15 \times 50 \times 3/6 = 375$) 个待分析的 ERP 波形.图 1 为随机抽取的一名说谎者和一名诚实者在 Pz 电极上的不同平均次数下的 ERP 波形(0 时刻代表刺激开始时间),可以看出,随着平均次数的增多,说谎者的 ERP 波形特征越来越明显,而诚实者的 ERP 波形特征基本不变.

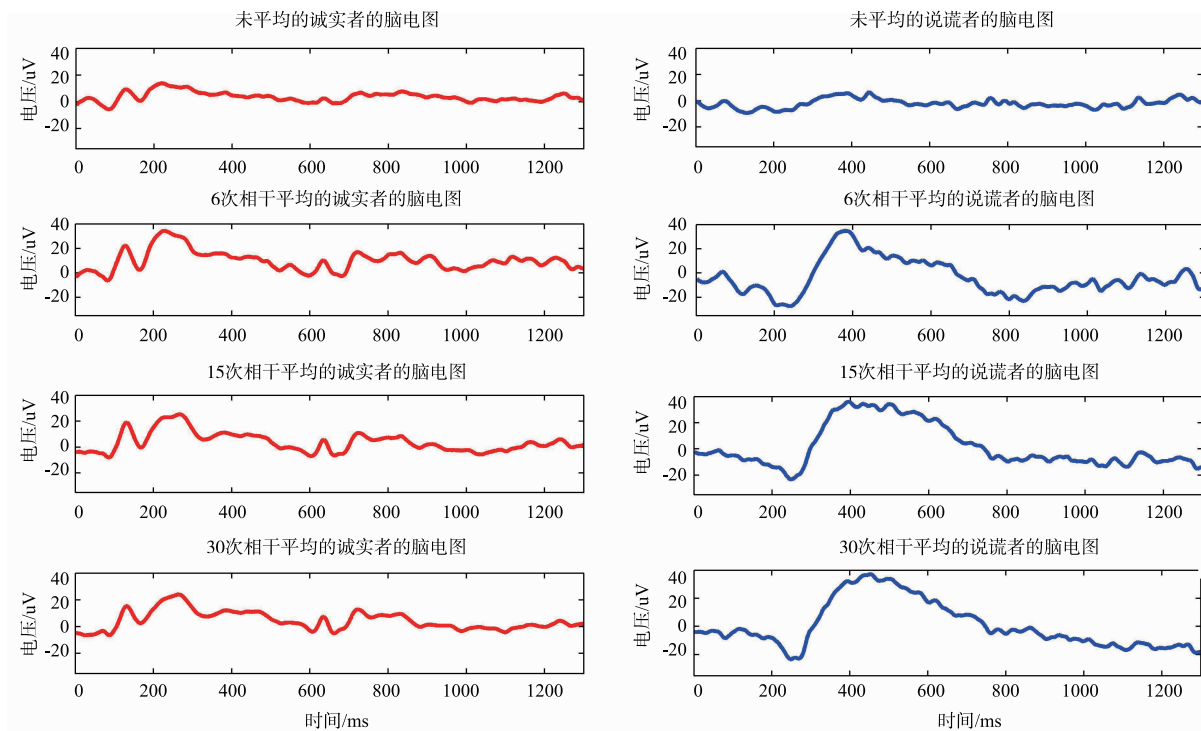


图1 两名受试者(左列为诚实者,右列为说谎者)在不同平均次数下的ERP波形

3.4 特征提取

将 6 次相干平均后的单一导联脑电信号作为样本熵算法的输入信号,分别取相邻的 m 点构成 (m 维) 矢量序列,通过计算样本熵观察维数由 m 增加到 $m+1$ 时,序列产生新模式的可能性大小.样本熵的值与嵌入维数 m 和相似性容限 r 的取值有关. m 值越大,计算的数据量越大,耗费时间越长. r 值越小噪声对结果的影响越显著,反之亦然.根据以往研究学者的大量数据统计,当 m 取 2 或 3, r 取 0.1 ~ 0.25 时,计算得到的样本熵具有较好的统计特性,此时诚实和说谎熵值区分度

较大,综合考虑本研究取 $r=0.2$.

由于 P300 信号主要产生于刺激后 700ms 的时间段内,为更好的评价样本熵对于测谎研究的脑电信号分析的效果,将 P300 信号响应时间段分为 4 种不同的时间段进行分析,定义 A 代表各时间段变量.从刺激后的 700ms 的信号中,将不同时间段的 EEG 信号取出分别进行熵值计算.

4 实验结果与分析

通过计算得出以下结果:处于诚实状态时 EEG 信

号的熵值波动范围(方差)明显小于说谎状态下的波动范围;说谎时 EEG 信号的熵值高于诚实时的熵值. 对两类受试者的熵值用独立样本 t 检验进行统计分析,得出在不同 A 及 m 值组合下,两类受试者的样本熵均具有显著差异($p < 0.05$).

对于本文提出的基于样本熵的特征提取方法,实质为二值分类问题. 为有效区分两类受试者,必须提出一个合理的样本熵阈值. 本文使用受试者工作特征曲线(Receiver Operating Characteristic curve, ROC 曲线)分析方法^[11,17,18]对阈值进行选取. 由于过长数据量的脑电

信号采集容易受到外界因素干扰,且样本熵只需较短的数据就能达到有效分析的目的,因此本文 A 选择 400ms, 500ms, 600ms, 650ms 四个时间长度进行分析. 图 2 显示了在不同 m 和 A 值组合下的 ROC 曲线. 每一列代表不同的 A 值,每一行代表不同的 m 值. 可以看出,当 $m=3, A=650$ ms 时,ROC 曲线的曲线下面积(Area Under Curve, AUC)最大,通过对比得出当样本熵的阈值为 0.453 时,平均分类准确率最高(敏感度 = 92.70%, 特异度 = 92.6%).

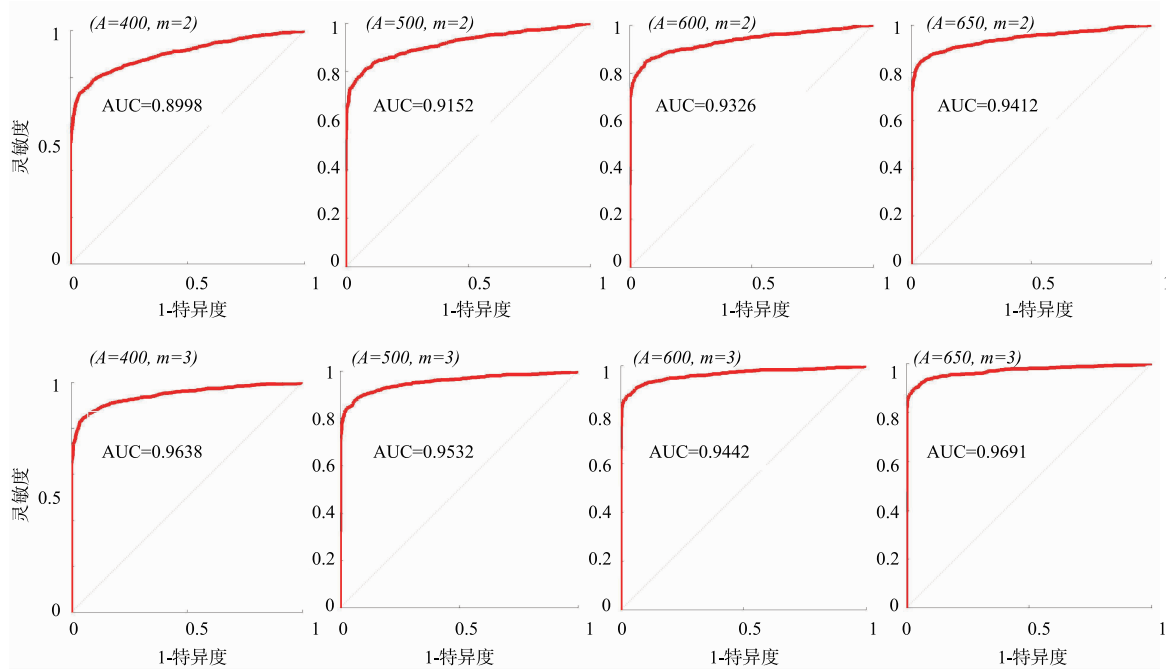


图2 不同 m 和 A 值得组合下的 ROC 曲线

表 1 不同 m 和 A 值的组合下两类受试者的分类准确率结果

| A 及 m 值组合 | 准确率 | | 样本熵 阈值 | ROC 曲线的 AUC 值 |
|------------------|-------|-------|-----------|------------------|
| | 敏感度 | 特异度 | | |
| (A = 400, m = 2) | 0.825 | 0.823 | 0.558 | 0.899 |
| (A = 400, m = 3) | 0.880 | 0.887 | 0.420 | 0.944 |
| (A = 500, m = 2) | 0.849 | 0.847 | 0.572 | 0.915 |
| (A = 500, m = 3) | 0.899 | 0.897 | 0.436 | 0.953 |
| (A = 600, m = 2) | 0.870 | 0.878 | 0.585 | 0.933 |
| (A = 600, m = 3) | 0.919 | 0.914 | 0.445 | 0.964 |
| (A = 650, m = 2) | 0.880 | 0.885 | 0.590 | 0.941 |
| (A = 650, m = 3) | 0.927 | 0.926 | 0.453 | 0.969 |

截止目前,虽然国内外已经有测谎准确率的报道,由于测谎方式(尤其本文是基于少次刺激)及测谎数据处理方式的不同,直接进行测谎准确率的对比并不合

适. 本文使用 BAD 和 BCD 这两种测谎方法对实验数据进行了分析,得出的测谎准确率与 Abootalebi 等作者的测谎准确率及本文所采取的测谎方法得出的结果进行了比较,结果显示在表 2 中.

表 2 不同测谎方法下的分类准确率结果

| 测谎方法 | 准确率 | | 平均 准确率 |
|---------------------------|-------|-------|-----------|
| | 敏感度 | 特异度 | |
| BAD | 0.792 | 0.831 | 0.740 |
| BCD | 0.811 | 0.864 | 0.802 |
| 小波特征提取方法 ^[11] | 0.730 | 0.820 | 0.790 |
| 多域联合特征提取方法 ^[5] | 0.670 | 0.880 | 0.861 |
| 本文提出方法 | 0.927 | 0.926 | 0.927 |

注释:多域联合特征提取方法指时域、频域及小波域特征联合.

5 结论

本文首次将非线性信号处理方法-样本熵用于测谎研究中的 EEG 信号分析. 通过对单个电极上 ERP 信号的熵值计算, 合理设置样本熵阈值, 得到 92.65% 的平均分类准确率. 样本熵是表征系统复杂度的重要非线性参数方法, 只需少量数据就能反映信号的非线性特征, 当模板匹配精度要求比较高时, 样本熵要明显优于现在广泛使用的近似熵.

本文研究过程中, 对原始脑电信号进行 6 次平均进行降噪处理. 实际测试过程中, 可只对被试者使用约 6 次刺激进行测试, 因此相对于 BCD 和 BAD 等方法, 本文提出少次刺激的测谎方法, 使得测谎过程更加简便, 大大降低了被测人员的疲劳程度, 反而有利于提高测谎准确率.

本文使用样本熵对说谎及诚实人的脑电信号进行分析, 计算结果表明样本熵适于测谎研究中脑电信号的分析, 因而样本熵的分析为测谎提供了一种新的途径, 另外, 其它熵的指标是否适合用于测谎研究, 这些都值得进一步研究.

本文对提取的 Pz 电极的脑电信号进行分析, 后期可以对其它的电极的脑电信号进行分析, 另外本文是通过先验经验对样本熵算法中的参数 r 进行取值, 将来还可以讨论不同的 r 值对测谎准确率的影响.

参考文献

- [1] Cui Q, E J Vanman, D Wei, et al. Detection of deception based on fMRI activation patterns underlying the production of a deceptive response and receiving feedback about the success of the deception after a mock murder crime [J]. *Social Cognitive & Affective Neuroscience*, 2014, 9 (10): 1472 - 1480.
- [2] Zhang Peng-yuan, Shao Jian, Yan Yong-hong. A two-step keyword spotting method using fuzzy search algorithm [J]. *Chinese Journal of Electronics*, 2008, 17 (1): 33 - 36.
- [3] 张丹丹, 罗跃嘉. ERP 的神经机制: 诱发模型还是相位重排模型 [J]. *心理科学进展*, 2011, 19(4): 487 - 498.
Zhang Dan-dan, Luo Yue-jia. Neural origin of erps: evoked model or phase reset model [J]. *Advances in Psychological Science*, 2011, 19(4): 487 - 498. (in Chinese)
- [4] Rosenfeld J P, Nasman V T, Whalen R, et al. Late vertex positivity in event-related potentials as a guilty knowledge indicator: a new method of lie detection [J]. *International Journal of Neuroscience*, 1987, 34 (1 - 2): 125 - 129.
- [5] Abootalebia V, Moradib M H, Khalilzadehc, et al. A new approach for EEG feature extraction in P300-based lie detection [J]. *Computer Methods & Programs in Biomedicine*, 2009, 94(1): 48 - 57.
- [6] Gao Jun-feng, Yang Yong, Zhang Wen-Jia, Tao Chun-yi, Guan Jin-an, Rao Ni-ni. A novel approach for lie detection based on F-score and extreme learning machine [J]. *Plos One*, 2013, 8(6): e64704.
- [7] 陈洪波, 李蓓蕾, 陈真诚. 基于 ICA 的脑电信号 P300 少次自动提取 [J]. *电子学报*, 2012, 40(6): 1257 - 1262.
Chen Hong-bo, Li Beilei, Chen Zhen-cheng. Automatically extract P300 within several trials from EEG based on ICA [J]. *Acta Electronica Sinica*, 2012, 40(6): 1257 - 1262. (in Chinese)
- [8] 程学梅, 崔园. 用 EEGLAB 分析脑电信号 [J]. *计算机与数字工程*, 2014, (10): 052.
Cheng Xue-mei, Cui Yuan. EEGLAB analysis of EEG signals [J]. *Computer and Digital Engineering*, 2014, (10): 052. (in Chinese.)
- [9] Labkovsky E, Rosenfeld J P. A novel dual probe complex trial protocol for detection of concealed information [J]. *Psychophysiology*, 2014, 51(11): 1122 - 1130.
- [10] Rosenfeld J P, Hu X, Pederson K. Deception awareness improves P300-based deception detection in concealed information tests [J]. *International Journal of Psychophysiology Official Journal of the International Organization of Psychophysiology*, 2012, 86(1): 114 - 121.
- [11] Abootalebi V, Moradi M H, Khalilzadeh M A. A comparison of methods for ERP assessment in a P300-based GKT [J]. *International Journal of Psychophysiology Official Journal of the International Organization of Psychophysiology*, 2006, 62(2): 309 - 320.
- [12] Zhao Q, Hu B, Shi Y, et al. Automatic identification and removal of ocular artifacts in EEG-improved adaptive predictor filtering for portable applications [J]. *IEEE Transactions on Nanobioscience*, 2014, 13(2): 109 - 117.
- [13] Lozano V A, Escandar G M. Application of multivariate empirical mode decomposition and sample entropy in eeg signals via artificial neural networks for interpreting depth of anesthesia [J]. *Entropy*, 2013, 15(9): 3325 - 3339.
- [14] 王攀, 沈继忠, 施锦河. 基于小波变换和时域能量熵的 P300 特征提取算法 [J]. *仪器仪表学报*, 2011, 32(6): 1284 - 1289.
Wang Pan, Shen Ji-zhong, Shi Jin-he. P300 feature extraction algorithm based on wavelet transform and temporal energy entropy [J]. *Chinese Journal of Scientific Instrument*, 2011, 32(6): 1284 - 1289. (in Chinese)
- [15] Gao Jun-feng, Lu Liang, Yang Yong, et al. A novel con-

cealed information test method based on independent component analysis and support vector machine [J]. *Clinical EEG and Neuroscience*, 2012, 43(1):54-63.

- [16] 顾凌云, 吕文志, 杨勇, 等. 基于 PCANet 和 SVM 的谎言测试研究 [J]. *电子学报*, 2016, 44(8):1969-1973.
Gu Ling-yun, Lv Wen-zhi, Yang Yong, et al. Deception-detection study based on PCANet and support vector machine [J]. *Acta Electronica Sinica*, 2016, 44(8):1969-1973. (in Chinese)
- [17] Gao Jun-feng, Tian Hong-jun, Yang Yong, et al. A novel

algorithm to enhance P300 in single trials: application to lie detection using F-score and SVM [J]. *Plos One*, 2014, 9(11):e109700.

- [18] 师夏阳, 马赛兰, 胡永健, 等. 一种基于像素块的纹理优先自适应隐写算法 [J]. *电子学报*, 2015, 43(6):1094-1100.
Shi Xia-yang, Ma Sai-lan, Hu Yong-jian, et al. A pixel block-based adaptive steganographic algorithm with embedding priority given to image textures [J]. *Acta Electronica Sinica*, 2015, 43(6):1094-1100. (in Chinese)

作者简介



高军峰 男, 1973 年生于湖北武汉. 副教授, 硕士生导师, 2005 年和 2011 年分别在武汉理工大学和西安交通大学获得工学硕士和工学博士学位. 现为中南民族大学教师, 主要从事生物医学信号处理、神经网络和机器学习等方面的研究工作.

E-mail: junfengmst@163.com



司慧芳 女, 1993 年出生于河南济源, 硕士研究生, 研究方向为脑电信号处理与模式识别.



杨勇 (通讯作者) 男, 1976 年 10 月生于湖北鄂州. 2005 年博士毕业于西安交通大学生物医学工程专业, 现为江西财经大学信息管理学院教授. 主要研究方向为医学图像/信号处理、模式识别.

E-mail: greatyangy@126.com