

基于改进的符号转移熵的心脑电信号耦合研究*

吴莎¹⁾ 李锦²⁾ 张明丽³⁾ 王俊^{1)†}

1) (南京邮电大学, 图像处理与图像通信江苏省重点实验室, 南京 210003)

2) (陕西师范大学, 物理学与信息技术学院, 西安 710062)

3) (西安市第一医院, 干部病房消化专业, 西安 710002)

(2013年6月20日收到; 2013年8月19日收到修改稿)

试图探究动力系统耦合关系一直以来都是国内外众多学者关注的热点, 传统的时间序列符号化分析方法会使研究结果受序列非平稳性的严重影响, 本文在原有转移熵的研究基础上, 应用粗粒化提取, 经过理论与实验的分析, 发现心脑电信号耦合研究中的转移熵值在不同提取情况下对应不同的分布趋势, 并选择效果最好的信号数据提取方法用在之后的应用分析中. 此外, 对时间序列符号化方法提出改进, 采用动态的自适应分割方法. 实验结果表明, 无论清醒期还是睡眠期, 改进的符号转移熵算法观测分析到的心脑电信号耦合作用更显著, 能更好的捕捉到信号中的动态信息、系统动力学复杂性的改变, 更利于医学临床实践应用中的检测, 在分析非平稳的时间序列上具有更好的效果.

关键词: 心脑电信号, 粗粒化, 符号转移熵, 基本尺度

PACS: 87.85.-d, 05.45.-a

DOI: 10.7498/aps.62.238701

1 引言

对人体各种生物电的研究一直是人们了解各器官的功能、临床诊断与治疗的重要手段. 脑电 EEG(electroencephalogram) 信号是通过电极记录下的脑电细胞群自发性节律性电活动, 心电图 ECG(electrocardiogram) 反映了心脏兴奋的产生、传导和恢复过程中的电变化, 这些参量都与人体的生理病理状态密切相关. 医学临床实践中采用脑电、心电图进行检测和分析有着很重要的应用价值. 20世纪90年代开始, 非线性动力学方法被逐步应用于生理电信号的研究中, 并取得一定进展. 其中, 心脑电信号是非常复杂的非平稳信号, 相比生物电信号采集技术的进步, 其分析技术还有待进一步发展^[1-3], 目前主要的分析方法有时域频域分析^[4-9]、神经网络^[9-14]、混沌分析^[15-17]等, 上述方法代表了现代心脑电信号研究的方向.

从物理学范畴到神经科学研究, 自从同步和相

关的复杂交互现象被检测到之后, 试图探究动力系统耦合关系一直都是国内外众多学者关注的热点. 近年来, 相关领域的调查研究主要涉及两个方面, 即耦合强度与方向的检测与量化. 但是, 运动的潜在方式尚未清晰, 只能通过不断的实验对时间序列进行分析来获得具体的定量描述. 时间序列符号化分析是由符号动力学、混沌序列分析理论发展起来的一种方法^[18,19], 其实质是指把原始的时间序列转换为若干符号组成的序列, 再进一步分析. 传统的符号动力学方法多采用静态范围来划分符号区域, 符号化过程中, 原始序列会丢失部分细节信息, 虽然动力学特征保留了下来, 但由于序列的非平稳性, 结果还是会受到严重影响.

针对以上问题, 本文在原有符号转移熵^[18]研究的基础上, 对时间序列符号化方法提出改进, 采用了动态的自适应分割法, 对相空间中的每一个 m 维矢量, 计算其基本尺度来选择符号划分的标准. 通过对清醒期和睡眠期的心脑电信号应用该方法计算相互之间的符号转移熵进行耦合作用的量化,

* 国家自然科学基金(批准号: 61271082, 61201029, 61102094)和江苏省自然科学基金(批准号: BK2011759, BK2011565)资助的课题.

† 通讯作者. E-mail: wangj@njupt.edu.cn

综合分析耦合强度与量化方向等因素, 与传统算法所得结果比较, 该方法能更好的捕捉到信号中的动态信息以及动力学系统复杂性的改变, 在分析非平稳的时间序列上具有更好的效果.

2 转移熵及改进的符号转移熵

2.1 转移熵

熵 (entropy) 通常用来衡量系统的混乱程度, 在通信领域中表示不确定性, 在生命科学中表示活动的无序性.

相对熵库尔巴克散度

$$K_I = \sum_i p(i) \cdot \log \frac{p(i)}{q(i)}$$

是两个随机分布之间距离的度量. 对于单个状态 j , 我们有

$$K_j = \sum_i p(i|j) \cdot \log \frac{p(i|j)}{q(i|j)},$$

将 j 求和, 就有

$$K_{I|J} = \sum_{i,j} p(i,j) \cdot \log \frac{p(i|j)}{q(i|j)}. \quad (1)$$

对于有联合概率分布 $p_{IJ}(i,j)$ 的两个过程 I 和 J , 若两个系统相互独立, 便有

$$M_{IJ} = \sum p(i,j) \cdot \log \frac{p(i,j)}{p(i)p(j)}. \quad (2)$$

这时相应的库尔巴克熵就变成了著名的互信息公式. 可见, 用互信息来量化两个过程独立性的偏差情况, 是一种很自然的方式. 但是, 当交换 I 和 J 时, M_{IJ} 是对称的, 不包含任何方向的信息. 若在其中一个变量的计算上引入时间延迟, 那么互信息就可以在某种程度上获得一定的方向信息, 如

$$M_{IJ}(\tau) = \sum p(i_n, j_{n-\tau}) \cdot \log \frac{p(i_n, j_{n-\tau})}{p(i)p(j)}. \quad (3)$$

信源的熵率

$$\begin{aligned} H(i_{n+1}|i_n^{(k)}) \\ = h_I = - \sum p(i_{n+1}, i_n^{(k)}) \cdot \log p(i_{n+1}|i_n^{(k)}), \end{aligned} \quad (4)$$

其中, $i_n^{(k)} = (i_n, \dots, i_{n-k+1})$. 想要研究多个过程之间的动力学信息, 可以将熵率的应用推广到多个系统, 因为过程的动力学信息是包含在转移概率里的, 最直接的方式是通过将 h_I 推广到两个过程 (I, J) 来构

造互信息率. 相应的库尔巴克熵在交换 I 和 J 时仍然是对称的. 这就有

$$\begin{aligned} h_{IJ} = - \sum p(i_{n+1}, i_n^{(k)}, j_n^{(k)}) \\ \times \log p(i_{n+1}|i_n^{(k)}, j_n^{(k)}). \end{aligned} \quad (5)$$

当 I 和 J 相互独立时, 便有:

$$\begin{aligned} h'_{IJ} = - \sum p(i_{n+1}, i_n^{(k)}, j_n^{(k)}) \\ \times \log p(i_{n+1}|i_n^{(k)}). \end{aligned} \quad (6)$$

因此, 定义转移熵 TE (transfer entropy)

$$\begin{aligned} T_{J \rightarrow I} = h'_{IJ} - h_{IJ} = \sum p(i_{n+1}, i_n^{(k)}, j_n^{(k)}) \\ \times \log \frac{p(i_{n+1}|i_n^{(k)}, j_n^{(k)})}{p(i_{n+1}|i_n^{(k)})}. \end{aligned} \quad (7)$$

2.2 改进的符号转移熵

转移熵对参数有着较高的协调性要求, 对噪声也较为敏感, 于是又有了基于符号化技术的新办法, 即符号转移熵 STE (symbolic transfer entropy)^[18].

符号动力学中很重要的一步, 就是根据原始时间序列的量值来确定符号划分的范围, 以便将原始序列转化为符号序列. 传统的符号动力学分析方法采用静态划分, 在符号化的过程中, 不可避免丢失细节信息, 虽然动力学特征保留了下来, 但由于序列的非平稳性, 结果还是会受到严重影响.

本文采用一种动态的自适应分割方法, 时间序列 x , 长度为 N :

$\{x(i) : 1 \leq i \leq N\}$, 将序列 x 嵌入 m 维相空间,

$$X(i) = [x(i), x(i+\tau), \dots, x(i+(m-1)\tau)], \quad (8)$$

其中 τ 为延迟时间.

接下来, 对于每一个 m 维矢量 $X(i)$, 我们计算其所有相邻数据差值的方均根值, 称为基本尺度 BS (basic scale)^[20], 即

$$BS(i) = \sqrt{\frac{\sum_{j=1}^{m-1} (x(i+j) - x(i+j-1))^2}{m-1}}. \quad (9)$$

根据以上结果, 将 $\alpha \times BS$ 作为符号划分的标准, 由于每一个 m 维矢量的 BS 不同, 划分标准也因此是动态变化的. 于是每一个 m 维矢量就转换成 m 维的符号序列: $S_i(i) = \{s(i), s(i+1), \dots, s(i+m-1)\}$.

1)}, $s \in L, (L = 0, 1, 2, 3),$

$$S_i(i) = \begin{cases} 0, & \bar{x} < x_{i+k} \leq \bar{x} + \alpha \times BS, \\ 1, & x_{i+k} > \bar{x} + \alpha \times BS, \\ 2, & \bar{x} - \alpha \times BS < x_{i+k} \leq \bar{x}, \\ 3, & x_{i+k} \leq \bar{x} - \alpha \times BS, \end{cases} \quad (10)$$

其中 $i = 1, 2, 3, \dots, N - m + 1, k = 0, 1, 2, \dots, m - 1,$ \bar{x} 、 BS 分别表示第 i 个 m 维矢量的均值、基本尺度. α 是常量, 取值过大会丢失细节信息, 无法较好的捕捉信号中的动态信息; 取值过小, 会加重噪声的影响 [21].

3 粗粒化 (coarse graining) 提取

相空间中, 点是最小的细粒, 若将一组点粗略地当成一个态, 即粗粒化. 实验表明, 多尺度谱函数 $F(\tau, \alpha)$ 中, 改变尺度因子 τ 可对时间序列粗粒化, 实际上就是间接地改变生理信号的采样频率.

如果 I 是通过连续系统 X 以分辨率 r 粗粒化获得, 那么熵 $H_X(r)$ 和熵率 $h_X(r)$ 都将依赖于采样, 且当时 $r \rightarrow 0,$ 由于 $-\log r,$ 它们将发散. 对于确

定性动力学系统的特例, $\lim_{r \rightarrow 0} h_X(r) = h_{KS}$ 极限可能存在, 称为柯尔莫戈洛夫-希奈熵 (K-S 熵是在无限的尺度下 block 熵的极限, 可用来描绘系统的动态随机程度). 而就互信息 MI 来说, 情况相反: 对于噪声相互依赖的系统, $\lim_{r \rightarrow 0} M_{XY}(r)$ 有限且与采样无关; 对于确定性的耦合系统, 当 $r \rightarrow 0$ 时 $M_{XY}(r)$ 发散.

对于由连续系统 (X, Y) 粗粒化的状态 $(I, J),$ $\lim_{r \rightarrow 0} T_{Y \rightarrow X}(r)$ 极限存在, 且与采样无关; 对于确定性的耦合系统, 当 $r \rightarrow 0$ 时 $T_{Y \rightarrow X}(r)$ 发散. 从这方面来看, 转移熵与互信息有着一致的趋势.

本文根据受试者 slp41 采集到的多参数睡眠数据, 分别在 r 为 0.01, 0.05, 0.1, 0.5, 1 时提取清醒期和 NREM 睡眠 I 期的 1 导 EEG(C4-A1) 信号和 1 导 ECG 信号, 各 5 组, 计算 TE 值. 实验结果见表 1.

同理, 可得出转移熵 $T_{ECG \rightarrow EEG}$ 在不同 r 时的值, 根据结果可分别画出清醒期和 NREM 睡眠 I 期 5 组 TE 值在不同 r 时的分布趋势, 如图 1 所示.

由图可以看出, 不同 r 的 TE 值有着如上分析的趋势, 且当 r 为 0.5 时 TE 值最大, 效果最好, 故在之后的应用分析中可选择此种采样情况提取信号数据.

表 1 转移熵 $T_{EEG \rightarrow ECG}$ 在不同 r 时的实验结果. W: 清醒期, S: NREM 睡眠 I 期

| | r=0.01 | | r=0.05 | | r=0.1 | | r=0.5 | | r=1 | |
|-----|--------|-------|--------|-------|-------|-------|-------|-------|-------|-------|
| | W | S | W | S | W | S | W | S | W | S |
| 1 | 0.023 | 0.037 | 0.163 | 0.129 | 0.206 | 0.237 | 0.628 | 1.121 | 0.567 | 0.695 |
| 2 | 0.021 | 0.039 | 0.116 | 0.148 | 0.188 | 0.279 | 0.981 | 0.899 | 0.593 | 0.561 |
| 3 | 0.031 | 0.028 | 0.117 | 0.088 | 0.214 | 0.145 | 0.989 | 0.575 | 0.792 | 0.518 |
| 4 | 0.045 | 0.028 | 0.145 | 0.080 | 0.286 | 0.141 | 1.099 | 1.009 | 0.655 | 0.710 |
| 5 | 0.029 | 0.044 | 0.074 | 0.131 | 0.143 | 0.185 | 0.562 | 0.811 | 0.769 | 1.051 |
| 平均值 | 0.30 | 0.035 | 0.123 | 0.115 | 0.208 | 0.197 | 0.852 | 0.883 | 0.675 | 0.707 |
| 标准差 | 0.009 | 0.007 | 0.034 | 0.029 | 0.051 | 0.060 | 0.240 | 0.208 | 0.101 | 0.209 |

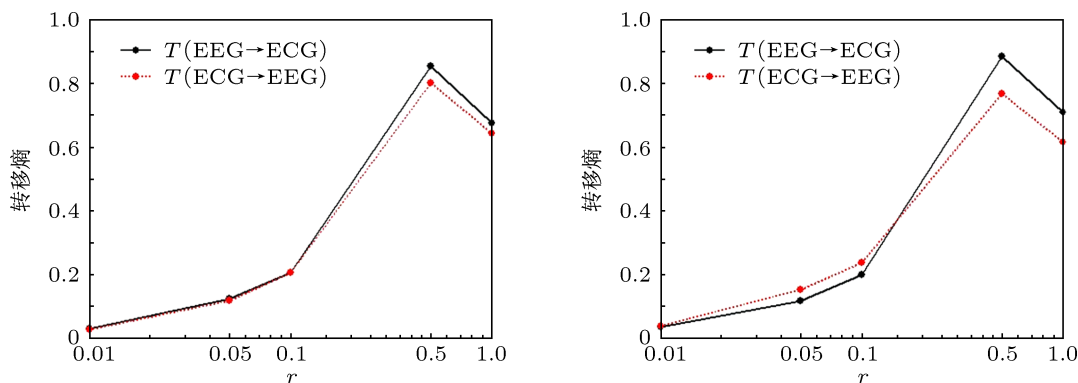


图 1 清醒期和 NREM 睡眠 I 期 5 组 TE 值在不同 r 时的分布趋势

4 改进的符号转移熵的心脑电信号睡眠分期耦合研究

4.1 实验数据

本文使用的清醒期、睡眠期 EEG, ECG 信号数据取自 PhysioBank 的 MIT-BIH Polysomnographic Database. 该库中记录的是多参数睡眠数据, 包含 1 导 EEG 信号、1 导 ECG、1 导 EMG(electromyographic signal)、1 导 EOG(eyectro-oculogram) 等多导睡眠信号, 数据基本采样频率为 250 Hz, 记录时长达 6 h, 附带以 30 s 为一个分期的注释信息.

4.2 实验方法

在以上研究的基础上, 我们首先对受试者 slp41 在清醒期和 NREM 睡眠 I 期分别提取 10 组长度为 5000 点的 1 导 EEG(C4-A1) 信号、1 导 ECG 信号, 采用 $r = 0.5$ 的粗粒化方法, 再生成 10 组数据, 对该 10 组数据分别利用传统的符号转移熵计算方法和本文提出的改进的符号化及编码算法进行处理, 计算出 TE 值, 从而分别获得清醒期和睡眠 I 期心脑电信号的耦合信息, 并作对比分析.

4.3 实验结果及分析

实验结果见表 2 和表 3.

将表 2 和表 3 的结果绘制成图, 如图 2 和图 3 所示.

从以上结果我们可以看出:

1) 改进的符号转移熵算法, 无论是清醒期还是睡眠期, $T_{EEG \rightarrow ECG}$ 的值总是大于 $T_{ECG \rightarrow EEG}$, 说明大脑到心脏的耦合作用更强; 此外, 无论 $T_{EEG \rightarrow ECG}$ 还是 $T_{ECG \rightarrow EEG}$, 清醒期的值均大于睡眠期, 说明心脑电信号的耦合作用在清醒期强于睡眠期, 与实际情况相符.

2) 与传统的符号动力学分析方法处理结果相比, 无论清醒期还是睡眠期, 改进的符号转移熵算法所得 $T_{EEG \rightarrow ECG}$, $T_{ECG \rightarrow EEG}$ 值均大于同等情况下传统算法计算的结果, 说明使用改进算法观测分析到的耦合作用更明显, 处理效果更显著.

3) 与传统的符号转移熵算法处理结果相比, 无论清醒期还是睡眠期, 改进的符号化及编码算法所得 $T_{EEG \rightarrow ECG}$, $T_{ECG \rightarrow EEG}$ 值在一定区间内波动范围

表 2 传统 STE 和改进 STE 算法在清醒期所得心脑电信号转移熵结果对比

| 信号 ($r = 0.5$) | 传统的 STE 算法 | | 改进的 STE 算法 | |
|------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| | $T_{EEG \rightarrow ECG}$ | $T_{ECG \rightarrow EEG}$ | $T_{EEG \rightarrow ECG}$ | $T_{ECG \rightarrow EEG}$ |
| 1 | 0.6959 | 0.5508 | 1.0112 | 0.9288 |
| 2 | 0.7636 | 0.9552 | 0.9727 | 0.8678 |
| 3 | 0.7002 | 0.6799 | 1.0898 | 1.1281 |
| 4 | 0.6669 | 0.6051 | 0.9638 | 1.0060 |
| 5 | 0.8929 | 0.4663 | 0.5131 | 0.5571 |
| 平均值 | 0.7439 | 0.6515 | 0.9101 | 0.8976 |
| 标准差 | 0.0905 | 0.1869 | 0.2274 | 0.2137 |

表 3 传统 STE 和改进 STE 算法在睡眠 I 期所得心脑电信号转移熵结果对比

| 信号 ($r = 0.5$) | 传统的 STE 算法 | | 改进的 STE 算法 | |
|------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| | $T_{EEG \rightarrow ECG}$ | $T_{ECG \rightarrow EEG}$ | $T_{EEG \rightarrow ECG}$ | $T_{ECG \rightarrow EEG}$ |
| 1 | 0.6531 | 0.9518 | 0.8805 | 1.1281 |
| 2 | 0.7054 | 0.5733 | 0.6753 | 0.6969 |
| 3 | 0.5527 | 0.5656 | 1.1032 | 0.8871 |
| 4 | 0.8184 | 0.5993 | 0.5388 | 0.5095 |
| 5 | 0.9028 | 0.4081 | 0.8456 | 0.6951 |
| 平均值 | 0.7265 | 0.6196 | 0.8087 | 0.7833 |
| 标准差 | 0.1374 | 0.2004 | 0.2144 | 0.2345 |

清醒期两种 STE 算法心脑电信号转移熵结果对比

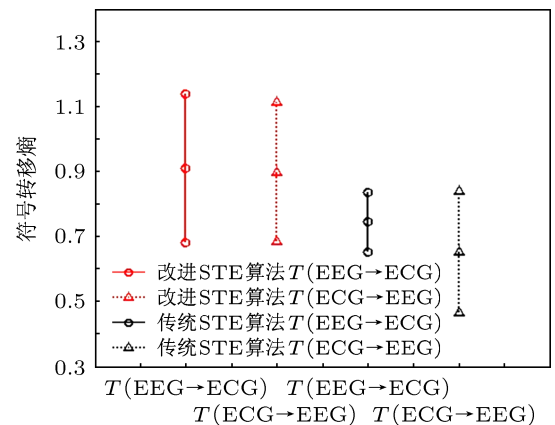


图 2 清醒期传统与改进的 STE 算法结果对比

睡眠期两种 STE 算法心脑电信号转移熵结果对比

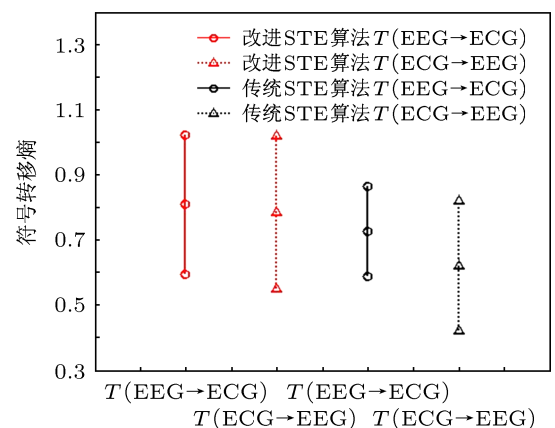


图 3 睡眠 I 期传统与改进的 STE 算法结果对比

更显著,且 $T_{\text{EEG} \rightarrow \text{ECG}}$ 与 $T_{\text{ECG} \rightarrow \text{EEG}}$ 的波动趋势更加一致和谐,能更好的捕捉信号中的动态信息,在分析非平稳时间序列上具有更好的效果.

5 结论

本文在原有转移熵的研究基础上,应用粗粒化提取,通过理论与实验的分析,发现在进行心脑电信号耦合研究时,转移熵 $T_{\text{EEG} \rightarrow \text{ECG}}$, $T_{\text{ECG} \rightarrow \text{EEG}}$ 的值在不同 r 时有着不同的分布趋势,并且当 r 为 0.5 时 TE 值最大,效果最好,故在之后的应用分析中选择此种采样情况提取信号数据.

传统的符号动力学方法采用静态划分,在符号化的过程中,不可避免丢失细节信息,虽然动力学特征保留了下来,但时间序列的非平稳性,还是使结果受到严重影响.本文在原有符号转移熵研究的基础上,对时间序列符号化方法提出改进,采用动态的自适应分割方法,对相空间中的每一个矢量,通过计算其基本尺度来划分范围.实验结果表明,无论清醒期还是睡眠期,改进的符号转移熵算法所得 $T_{\text{EEG} \rightarrow \text{ECG}}$, $T_{\text{ECG} \rightarrow \text{EEG}}$ 值均大于同等情况下传统算法计算的结果,说明使用改进算法观测分析到的耦合作用更显著,能更好的捕捉到信号中的动态信息,更利于医学临床实践等应用中的检测.

-
- [1] Fang X L, Jiang Z L 2007 *Acta Phys. Sin.* **56** 7330 (in Chinese) [方小玲, 姜宗来 2007 物理学报 **56** 7330]
- [2] Meng Q F, Zhou W, Chen Y H, Peng Y H 2010 *Acta Phys. Sin.* **59** 123 (in Chinese) [孟庆芳, 周卫东, 陈月辉, 彭玉华 2010 物理学报 **59** 123]
- [3] Ma Q L, Bian C H, Wang J 2010 *Acta Phys. Sin.* **59** 4480-4484 (in Chinese) [马千里, 卞春华, 王俊 2010 物理学报 **59** 4480]
- [4] Wang J, Ma Q L 2008 *Chin. Phys. B* **17** 4424
- [5] Hsu W Y 2012 *Int. J. Neural Syst.* **22** 51
- [6] Petrantonis P C, Hadjileontiadis L J 2012 *IEEE Ton Signal Proc.* **60** 2604
- [7] Thatcher R W 2012 *Dev. Neuropsychol.* **37** 476
- [8] Wang J, Yu Z F 2012 *Chin. Phys. B* **21** 018702
- [9] Musselman M, Djurdjanovic D 2012 *Expert Systems with Appl.* **39** 11413
- [10] Wang R F, Zhang J H, Zhang Y, Wang X Y 2012 *Biomed. Signal Proc. and Control* **7** 490
- [11] Orhan U, Hekim M, Ozer M 2012 *J. of Med. Syst.* **36** 2219
- [12] Acharya U R, Sree S V, Alvin A P C, Suri J S 2012 *Expert Syst. Appl.* **39** 9072
- [13] Siuly S, Li Y 2012 *IEEE T on Neur. Sys. and Reh. Eng.* **20** 526
- [14] Acharya U R, Molinari F, Sree S V, Chattopadhyay S, Ng K H, Suri J S 2012 *Biomed. Signal Proc. and Control* **7** 401
- [15] Rosenblum M G, Pikovsky A S 2001 *Phys. Rev. E* **64** 045202
- [16] Smirnov D A, Bezruchko B P 2003 *Phys. Rev. E* **68** 046209
- [17] Smirnov D A, Bodrov M B, Velazquez J L P, Wennberg R A, Bezruchko B P 2005 *Chaos* **15** 024102
- [18] Staniek M, Lehnertz K 2007 *Phys. Rev. Lett.* **99** 204101
- [19] Shen W, Wang J 2010 *Acta Phys. Sin.* **60** 118702 (in Chinese) [沈韡, 王俊 2010 物理学报 **60** 118702]
- [20] Huang X L, Cui S Z, Ning X B 2009 *Acta Phys. Sin.* **58** 8160 (in Chinese) [黄晓林, 崔胜忠, 宁新宝 2009 物理学报 **58** 8160]
- [21] Li J, Liu D Z 2012 *Acta Phys. Sin.* **61** 208701 (in Chinese) [李锦, 刘大钊 2012 物理学报 **61** 208701]

Coupling analysis of electrocardiogram and electroencephalogram based on improved symbolic transfer entropy*

Wu Sha¹⁾ Li Jin²⁾ Zhang Ming-Li³⁾ Wang Jun^{1)†}

1) (*Image Processing and Image Communications Key Lab., Nanjing Univ. of Posts & Telecomm., Nanjing 210003, China*)

2) (*College of Physics and Information Technology, Shaanxi Normal University, Xi'an 710062, China*)

3) (*Department of Gastroenterology, The No. 1 Hospital of Xi'an, Fen Rd 30, the South Street, Xi'an 710002, China*)

(Received 20 June 2013; revised manuscript received 19 August 2013)

Abstract

Exploration of the coupling relationship in dynamical system has always been a hot topic of many scholars at home and abroad, the traditional symbolic dynamics analysis method may lead to the results from the serious effect of non-stationary time series. This paper employs coarse graining extraction based on research of original transfer entropy. Through theoretical and experimental analysis, we find that the results of transfer entropy have different distribution trend under different extraction conditions in the coupling analysis of electroencephalogram and electrocardiogram. We choose the best effect of signal data extraction method and apply it to the later application analysis. Furthermore, this paper proposes improvement on the method of time series symbolization, using dynamic adaptive segmentation method. The experimental results show that the whether waking period or sleeping stage, coupling between electroencephalogram and electrocardiogram is more significant when using improved symbolic transfer entropy algorithm. It is also better to capture the dynamic information of the signal and the change of complexity of system dynamics, which is more conducive to clinical testing in practical application and has a better effect on the analysis of non-stationary time series.

Keywords: electrocardiogram and electroencephalogram, coarse graining, symbolic transfer entropy, basic scale

PACS: 87.85.-d, 05.45.-a

DOI: 10.7498/aps.62.238701

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61271082, 61201029, 61102094), and the Natural Science Foundation of Jiangsu Province (Grant Nos. BK2011759, BK2011565).

† Corresponding author. E-mail: wangj@njupt.edu.cn