

脑电信号多重粗粒化复杂度分析方法研究

肖毅, 陈善广, 韩东旭, 王春慧

(中国航天员科研训练中心人因工程重点实验室 北京 海淀区 100194)

【摘要】为了准确地描述脑电信号的精细结构和复杂成分,提出了基于多重粗粒化的脑电信号复杂度方法。根据脑电信号幅值微弱,幅值跨越大,一定幅值范围的脑电具有特定认知和生理意义的脑电特性,采用多重赋值,可尽量保留对信号复杂结构的描述;基于统计学特点(以百分比为阈值)对已经归一化的信号划分幅值域,进行粗粒化,基于传统方法计算复杂度。该方法有效地克服了经典LZC复杂度不能描述脑电精细结构的不足。提出了当脑电信号在粗粒化后出现局部周期,且达到一定阈值后(如beta波或具有相似幅值特性的脑电信号达到一定阈值),4重复复杂度值小于10重复复杂度值的假设,初步建立了脑电节律与多重LZC的联系。仿真计算和实际数据验证表明,多重LZC方法结果更准确、合理,而且能反映脑电节律成分特性。

关键词 假设; 二重粗粒化; 精细结构; LZC复杂度; 多重粗粒化

中图分类号 TP391

文献标志码 A

doi:10.3969/j.issn.1001-0548.2013.03.029

EEG Complexity Method Based on Multiple Coarse-Grained Sequences

XIAO Yi, CHEN Shan-guang, HAN Dong-xu, and WANG Chun-hui

(National Key Laboratory of Science and Technology on Human Factors Engineering, Astronaut Training and Research Center Haidian Beijing 100194)

Abstract In order to describe the fine structure and complexity of EEG components accurately, a EEG complexity method is proposed based on multiple coarse-grained EEG components. The coarse-grained thresholds of the method are selected based on the normalized statistics characters of the EEG. The complexity value is computed according to the formula of the classical LZC method. The simulation results based on the logistic map validate that the multiple LZC is much better than conventional methods and the trend of the LZC is much closer to the Lyapunov exponents. The method overcomes the defect of classical LZC complexity method, which cannot describe the fine structure of EEG characteristics effectively. Finally an assumption is proposed based on the results of the BCI IV EEG data, that is, the quadruple LZC is less than the decuple LZC when the EEG was partial periodicity and the period reached a certain threshold. The assumption is validated by the simulation data based on the EEG characters, a relationship between EEG rhythms and multiple LZC is established by which.

Key words assumption; decuple; fine structure; LZC complexity; multiple coarse-grained

直观意义而言, LZC复杂度表征了一个时间序列出现新模式的概率, LZC值越大表明出现新模式的概率越大,复杂度越大^[1-3]。由于LZC复杂度物理意义明确,计算速度更快,在脑电信号处理方面得到了广泛的应用,目前已经成为探索脑功能和诊断疾病的又一大非线性动力学途径^[4-15]。文献[16]对复杂度进行了分析。

文献[17]等针对现有的粗粒化过程容易丢失重要的信息,提出了 C_0 复杂度,认为复杂运动是由规则运动和随机运动混合而成的,随机运动所占的分

额就是 C_0 复杂性描述的基础。然而,脑电信号本质上是确定信号(混沌信号也是确定信号,只是周期无穷大),反映了具体的认知过程和活动;随机的信号是噪声。其产生往往不是单一的神经元放电,而是多个神经元及其突触的共同作用形成的,因此,脑电信号可以理解为由多个产生函数的非线性组合而成,可见 C_0 复杂度的前提假设值得商酌。

针对目前应用的基于二重粗粒化赋值LZC复杂度不能充分描述脑电信号的复杂结构,造成信息丢失,结果的可靠性受到质疑。本文改进了LZC复杂

收稿日期: 2011-01-18; 修回日期: 2013-03-21

基金项目: 国家973计划(2011CB711000); 中国航天医学工程预先研究项目(2010SY5405010); 国家自然科学基金青年基金(71201148); 部级资助项目(HF2011ZB02, HF2011ZZA01)

作者简介: 肖毅 (1981-), 男, 博士, 主要从事脑电信号处理、脑力负荷、认知建模、人机环境系统工程等方面的研究。

度的粗粒化赋值方法,提高了计算结果的可靠性。

1 多重粗粒化复杂度方法

1.1 传统LZC复杂度算法的缺点

基于二重粗粒化的脑电信号复杂度不能充分描述脑电信号特征^[17]的原因在于:1)信号幅值非常微弱,幅值以 μV 计,以均值为阈值,将更微弱的信号与均值信号等同,缺少理论依据。(假设脑电最大幅值为 $100\ \mu\text{V}$,均值为 $50\ \mu\text{V}$,低于 $50\ \mu\text{V}$ 的脑电信号粗粒化为0,则 $49\ \mu\text{V}$ 与 $1\ \mu\text{V}$ 的脑电相同,明显存在缺陷)。2)信号幅值跨越较大,仅二值化,对脑电信号特征的描述过于简单,无法有效地反映脑电幅值的复杂特性。3)脑电是脑生理活动的反映,具有重要的生理意义和认知意义,特定的脑电信号幅值域可能具有一定的认知意义。

1.2 多重粗粒化复杂度方法的基本思想

基于多重粗粒化计算的复杂度,简称多重复杂度。对脑电信号而言,幅值是非常重要的物理指标,但不是所有的幅值信号都具有特定的意义,一定幅值范围的信号可能具有特定的意义,本文定义这种具有特定意义的幅值范围为幅值域。多重复杂度根据不同波段脑电信号粗略对应的幅值范围划分不同的幅值域,进行粗粒化,以反应不同波段对应的生理意义。

下文分析4重粗粒化和10重复复杂度方法。

1.3 方法步骤

多重复杂度方法从计算过程而言,基本遵循经典LZC^[2]复杂度方法的计算步骤,主要不同在于重数的选择及粗粒化过程,详细分析见2.4节。为了分析结果的准确性,尤其为了克服个体差异,算法还增加了其他的步骤,具体算法步骤如下:

1)数据预处理,主要是滤波,这是脑电信号处理中的必需步骤;

2)数据归一化,去除个体差异,确保信号分析结果不只受幅值影响;

3)多重粗粒化过程,确定粗粒化的重数与粗粒化方法,对预处理后信号进行多重粗粒化赋值(详见2.4节);

4)统计不同字串的数量 $c(n)$;

5)归一化,得到多重粗粒化LZC值。

1.4 多重参数和粗粒化方法选择

重数的选择需要考虑计算结果的精确性和运算量之间的平衡,一般而言,重数越多结果越精确,

计算量增加,计算耗时长。本文重数分别选取了4和10。本文的仿真数据表明,重数为10时,计算时间和所需的数据长度都增加。

1) 粗粒化方法

由于信号幅值已经归一化,可基于统计学特点确定粗粒化方法,本文以百分比为阈值,划分幅值域,进行粗粒化。具体的做法如下:

根据重数对滤波和归一化处理后的数据按照等分的原则进行粗粒化。如10重粗粒化,将数据均等的分为10个区间,分别为: $[0, 0.1]$ 、 $[0.1, 0.2]$ 、 $[0.2, 0.3]$... $[0.9, 1.0]$,粗粒化幅值分别为: -1 、 $+1$ 、 -2 、 $+2$ 、 -3 、 $+3$ 、 -4 、 $+4$ 、 -5 、 $+5$,根据各节律的幅值范围可知beta波对应区间为 $[0, 0.1]$ 、 $[0.1, 0.2]$ 粗粒化后值为 -1 、 $+1$ 、而 $[0.6, 0.7]$ 及以上区间对应的则为theta和delta波。4重粗粒化则将数据均等地分为4个区间分别为: $[0, 0.25]$ 、 $[0.25, 0.50]$ 、 $[0.50, 0.75]$ 、 $[0.75, 1.00]$,粗粒化幅值分别对应1~4,其中 $[0, 0.25]$ 对应节律为beta。

2) $b(n)$ 的确定

依据文献[1]可得, $\lim_{n \rightarrow \infty} c(n) = b(n) = n / \log_N n$,确定 $b(n)$,其中 N 为重数。

2 仿真分析

2.1 Logistic验证算法有效性

Logistic映射方程如下: $x = rx(1-x)$; $r = 3 \sim 4$;其吸引因子如图1a所示,长度为3 000点。

参数适当时,Logistic映射表现出混沌特性,而在此之前为周期性分叉和倍周期^[3]阶段。Logistic在周期性分叉阶段是周期信号,复杂度值很小,进入混沌阶段后,复杂度值增大,但应不大于1。

在图1a中,Logistic吸引因子在 $3 < \lambda < 3.45$ 为倍周期阶段,在 $3.45 < \lambda < 3.58$ 为周期性分叉阶段,在 $\lambda \geq 3.58$ 为混沌状态。在图1b中,Lyapunov指数的变化趋势如下:在数据区间 $[580, +\infty)$ ($\lambda > 3.58$)进入了快速变化复杂阶段即混沌状态;而以450点(附近)为中轴,右侧止于580点附近($3.45 < \lambda < 3.58$),其中左侧对称区域为周期变化,这与Logistic吸引子的变化类似,在数据点240点附近($\lambda \approx 3.24$)为中轴的左右区域也为周期性变化,符合Logistic的变化过程。图1c所示的最大复杂度值大于1,而图1d和图1e所示的复杂度值都小于1。由复杂度理论可知,图1d、1e所示的计算结果更为可靠。下面具体分析复杂度值与Lyapunov指数的相互联系。

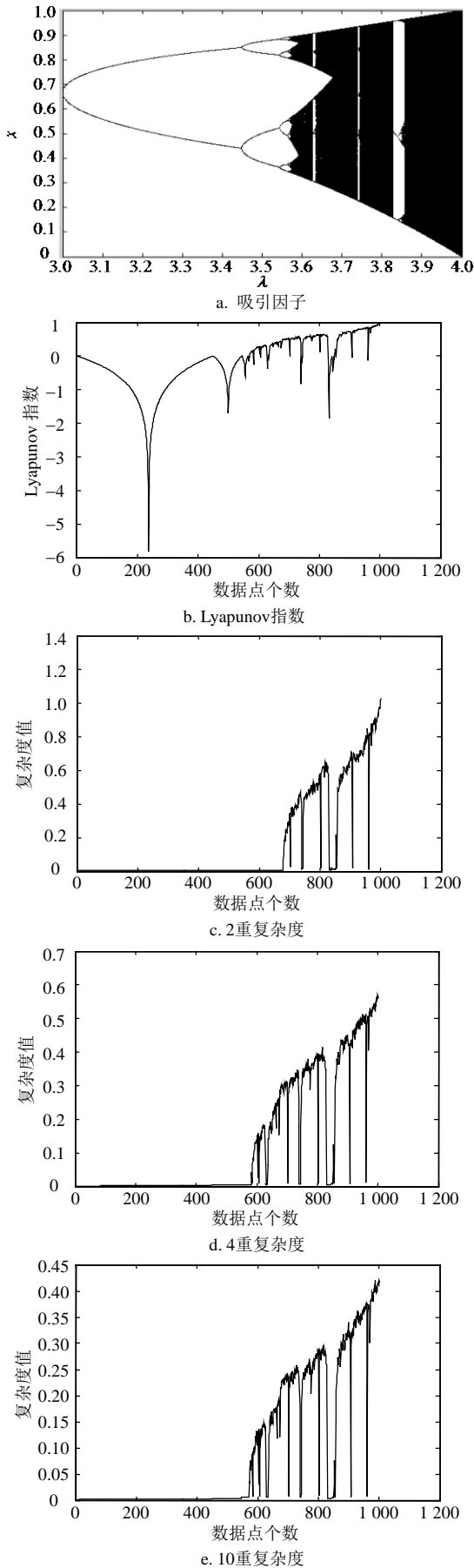


图1 Logistic映射及Lyapunov指数和复杂度

在图1c中, 2重复复杂度为两个阶段:

- 1) 数据区间[0, 680]为第一阶段, 复杂度值很小, 反映了Logistic的周期性特性;
- 2) 数据区间[680, +∞]为第二阶段, 复杂度值进入快速复杂变化期。与Lyapunov指数变化趋势比对可知, 二重复复杂度虽然能描述Logistic映射的大体变化趋势, 但无法刻画Logistic映射的复杂性和精细成分。

在图1e中, 10重复复杂度可精细地分为5个阶段:

- 1) 第一阶段为0及其附近很小的一段数据点, 复杂度值无穷接近0;
- 2) 第二阶段为此段之后到460点数据之间, 复杂度大于前一阶段值, 但依然非常小;
- 3) 数据区间[460, 560]为第三阶段;
- 4) [560, 580]为第四阶段, 复杂度稍大于前一阶段值, 但依然非常小;
- 5) [580, +∞]为第五阶段, 复杂度值进入快速复杂变化期。其中前四阶段复杂度值很小, 反映了Logistic映射的周期性特性, 第五阶段Logistic映射已经进入混沌状态, 这与Lyapunov指数变化趋势极为接近。由此可见, 10重复复杂度对数据的特性刻画更精细和准确。

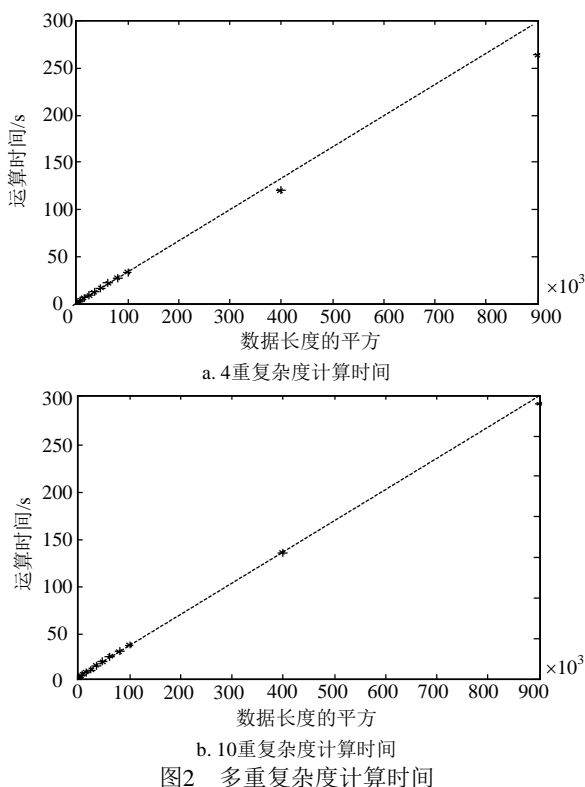
由图1可得到以下的结论:

- 1) 当Logistic为周期性分叉阶段时, Lyapunov指数变化平缓, 复杂度值很小, 且4重和10重复复杂度值接近0。
- 2) 多重复杂度与Lyapunov指数一样, 可反映Logistic周期性分叉特性, 图1e中10重复复杂度的变化趋势与Lyapunov指数几乎完全一致。
- 3) 在混沌阶段, 多重复杂度值具有与Lyapunov指数相同的变化趋势, 且当Lyapunov指数为0时, 复杂度也迅速下降, 与理论分析一致。
- 4) Logistic复杂度最大值小于1。

综上所述, 多重复杂度比2重复复杂度更能反映Logistic映射吸引因子的复杂程度和精细特性; 而10重比4重复复杂度对精细成分的描述更全面, 更好地描述了数据的精细特征。

2.2 计算长度和时间

依据文献[1]可知, 选择不同长度的随机数据分别计算复杂度, 当值接近1, 认为数据长度范围为计算所需长度(下文以DLength代替); 同时考察计算时间(下文以CTime代替)与所需数据长度的关系, 计算结果如图2所示。



2.2.1 计算长度

1) 4重复复杂度计算长度

① 当 $DLength < 10^4$ 时, 随机序列复杂度值非常接近且小于 $1(2 \times 10^3 \text{ 点值大于 } 1)$, 与理论分析一致, 表明该计算方法符合理论, 算法有效、可靠;

② 当 $DLength > 10^3$ 时, 复杂度值相对稳定, 且已大于 0.99 , 因此, 计算长度可选择 10^3 。

2) 10重复复杂度计算长度

① 当 $DLength > 10^3$ 时, 随机序列复杂度值大于 0.9 小于 1 , 可认为当 $DLength$ 趋于无穷大时, 随机序列复杂度将无限趋近于 1 , 与理论分析一致, 表明该计算方法符合理论, 有效、可靠。由于当 $DLength$ 为无穷时, 其复杂度值最接近 1 ;

② 当 $DLength > 10^4$ 时, 复杂度值相对稳定于 0.94 和 0.95 之间;

综上所述, 当 $DLength > 10^3$, 复杂度值已大于 0.9 , 在计算过程中可选择 10^3 , 既避免耗时过多, 又也可保证数据计算结果的可比性和有效性。

2.2.2 计算时间

由图2可见, 计算时间基本与数据长度的平方成线性关系, 与文献[3]的分析结果一致。

1) 4重复复杂度的计算时间

当 $DLength > 1 \times 10^4$, $CTime > 1 \text{ min}$; 当 $DLength > 2 \times 10^4$, $CTime > 2 \text{ min}$; 当 $DLength < 5 \times 10^3$, $CTime < 10 \text{ s}$ 。从计算时间而言, $DLength$ 不宜大于 5×10^3 。

2) 10重复复杂度的计算时间

当 $DLength > 1 \times 10^4$, $CTime > 1 \text{ min}$; 当 $DLength > 2 \times 10^4$, $CTime > 2.5 \text{ min}$; 当 $DLength < 5 \times 10^3$, $CTime < 10 \text{ s}$ 。从 $CTime$ 而言, $DLength$ 不宜大于 5×10^3 。

综上, $DLength$ 不宜大于 5×10^3 。

3 脑电数据验证和假设提出

本文采用视觉作业(4个不同的视觉认知任务)下采集的脑电数据进行验证, 计算了2重、4重和10重复复杂度。结果见表1。由表1可知, 被试在执行不同的认知任务时, 脑电信号具有不同的复杂度值, 表明不同任务下不同幅值脑电信号的排列复杂性存在差异。下面分析4重复复杂度值比10重复复杂度值小的原因, 其中不同节律脑波幅值及频率关系见表2。

表1 2重复复杂度计算结果

认知任务	数据区间	LZC2	LZC4	LZC10
3a任务2	(51 600 52 599)	0.698	0.478	0.603
3a任务3	(72 100 73 099)	0.548	0.219	0.384
3a任务4	(93 000 93 999)	0.598	0.234	0.405
3a任务5	(113 500 114 499)	0.568	0.503	0.591

表2 不同节律脑波幅值及频率对应关系

节律类型	频率范围/Hz	幅值范围/ μV
delta	1	20~200
theta	4~8	20~100
alpha	8~14	20~60
beta	14~30	2~20

假设信号经粗粒化后存在局部周期, 复杂度值减小, 且当局部周期信号的长度达到特定阈值时, 4重复复杂度值可能小于10重复复杂度值, 主要是因为新模式出现的次数减少所致。这表明在一段具有脑电幅频特性的周期信号内, 4重粗粒化后数据的局部周期性强于10重粗粒化后的局部周期性。

分析脑电节律信号的幅值特性可知, 当某时间段beta波(或具有相似幅值特性的脑波)处于优势时, 根据4重粗粒化的划分原则, 幅值为 1 , 10重粗粒化则为 $-1, 1$ 。即10重粗粒化中出现新模式(NP)的概率和次数增加, 并且大于特定阈值后, 则 $NP_4 / NP_{10} < \lg 4$, 即10重复复杂度值大于4重复复杂度。据此推测, 如果在某一时间段内脑电信号以beta波(或具有相似幅值特性的脑波)为主, 则该段脑电信号复杂度值减小, 且4重复复杂度值小于10重复复杂度值; 反之如果在某一时间段内脑电信号以theta波(或具有相似幅值特性的脑波)为主, 则该段脑电信号复杂度值增加,

且如果幅值在20~100 μV 不规则分布,可能4重复复杂度值大于10重复复杂度值。

为进一步分析上述假设,构造3组介于[0,1]长度为1 000的仿真数据,虽然绝对幅值与真实脑电数据不一致,但各种节律信号幅值的比例与真实脑电数据中的比例一致。计算结果见表3, LZCX表示X重复复杂度, NPX表示X重新模式, NP为新模式数目。

构造数据形式如下: x₁第一段300个数据点幅值介于0.01~0.2,中间一部分数据(400点)介于0.20~0.60,最后部分数据(300点)介于0.60~1.00,复杂度计算结果见表3中的x₁结果。x₂有300个数据点幅值介于0.01~0.2,一部分数据(400点)介于0.20~0.60,另一部分数据(300点)介于0.60~1.00,复杂度计算结果见表3中的x₂结果。数据按照(beta、alpha、theta、alpha、beta)顺序排列。x₃有200个数据点幅值介于0.01~0.2,一部分数据(450点)介于0.20~0.60,另一部分数据(350点)介于0.60~1.00,复杂度计算结果见表3中的x₃结果。数据按照(beta、alpha、theta、alpha、theta)顺序排列。

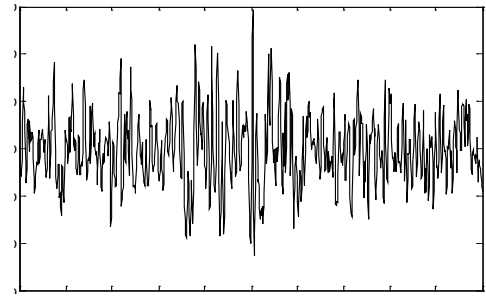
表3 仿真数据计算结果

数据	NP2	LZC4	NP10	LZC10	NP2	LZC2
x ₁	96	0.478	161	0.483	51	0.508
x ₂	98	0.488	164	0.492	54	0.538
x ₃	109	0.657	180	0.653	28	0.338

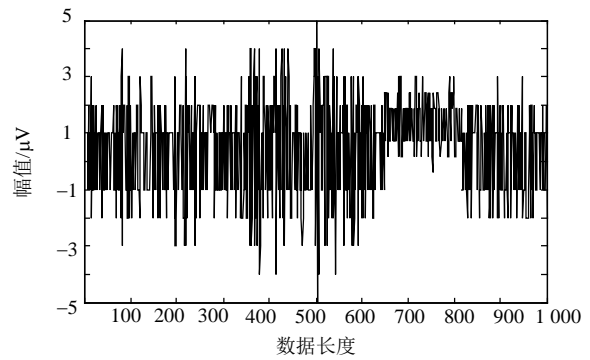
通过上述仿真数据分析可知,当beta连续数据点个数介于200~300时,4重复复杂度值大于10重复复杂度值;当beta连续数据点个数大于300,并且后续数据也具有一定规则时,10重复复杂度值大于4重复复杂度值。如果后续数据排列不规则(如不按x₁、x₂、x₃顺序排列),即使beta连续数据点个数足够,如900点,10重复复杂度值仍然小于4重复复杂度值。

复杂度值的变化与脑电节律信号的分布密切相关,理想情况下当beta波增加,同时其他节律信号的幅值分布也具有一定规则的情况下,4重复复杂度可能小于10重复复杂度。beta波的增加可能与多重复杂度值的下降相关。由图3可见,大量数据位于0~20 μV 之间,经归一化和粗粒化后,4重粗粒化后的周期性强于10重粗粒化后的周期性,幅值分布单一且趋于较小值,上述情况下4重复复杂度值小于10重复复杂度值。因此,上述假设具有理论可行性,且与实际有一定的吻合。基于本文的假设,可通过分析复杂度值的变化尤其是4重复复杂度值与10重复复杂度值的相互关系,进而判断脑电节律信号的排列特性,分析产生变化的原因,初步建立多重复杂度与脑电节律的联

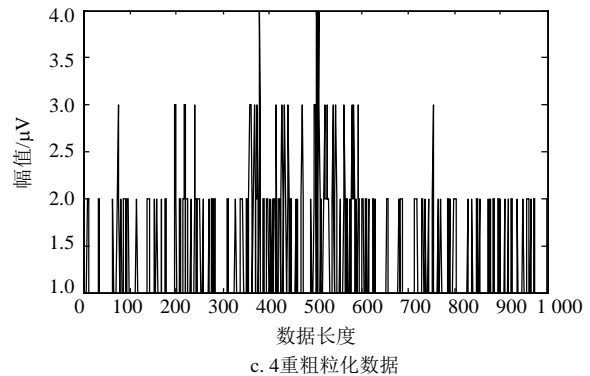
系,进而推断相应的认知特性。



a. 原始数据



b. 10重粗粒化数据



c. 4重粗粒化数据

图3 受试者3a任务2一段数据其粗粒化后幅值分布

4 结论

本文的研究表明,4重和10重复复杂度方法与2重复复杂度方法相比,具有计算结果更可靠、结果更精确、对信号精细结构的描述更充分的特点。

对计算结果的精确度要求较高,可选择10重复复杂度方法;对精确度的要求可适当降低,但对计算速度有较高要求,则可选择4重复复杂度方法。从分析的准确性而言,最好是两种方法都做计算,并分析4重与10重复复杂度值的变化关系。总之如何选择计算重数,应根据实际需要确定。

本文基于理论分析和数据验证研究了影响多重复杂度值的原因,发现多重复杂度值与脑电节律的分布密切相关,而且beta波的增加与多重复杂度值的

下降可能存在较为直接的关系:当beta波(或具有beta波幅值特性的信号)连续增加到一定数据长度,而且其他alpha、theta波(或具有其波幅特性)的分布具有一定规则时,4重粗粒化复杂度值小于10重复复杂度值。可能还有其他原因导致这些现象的产生,相关研究将在后续工作中进行。

参 考 文 献

- [1] LEMPEL A, ZIV J. On the complexity of infinite sequences[J]. *IEEE Trans in Theory*, 1976, 22(1): 75-81.
- [2] 李颖洁, 邱意弘, 朱贻盛. 脑电信号分析方法及其应用[M]. 北京: 科学出版社, 2009.
- LI Ying-jie, QIU Yi-hong, ZHU Yi-sheng. EEG signal analysis method and its application[M]. Beijing: Science Press, 2009.
- [3] KASPAR F, SCHUSTER H G. Easily calculable measure for the complexity of spatiotemporal patterns[J]. *Phys Rev A*, 1987, 36(2): 842-848.
- [4] 张泾周, 马颖颖, 李婷, 等. 基于复杂性测度的睡眠脑电分期处理方法研究[J]. *中国生物医学工程学报*, 2009, 28(3): 367-371.
- ZHANG Jing-zhou, MA Ying-ying, LI Ting, et al. A sleep EEG segmentation process based on complexity measure[J]. *Chinese Journal of Biomedical Engineering*, 2009, 28(3): 367-371.
- [5] 肖毅, 陈善广, 王春慧. 基于熵理论和复杂度的肌电信号分析[J]. *科技导报*, 2009, 27(21): 77-81.
- XIAO Yi, CHEN Shan-guang, WANG Chun-hui. Electromyography analysis based on the entropy theory and complexity measures[J]. *Science & Technology Review*, 2009, 27(21): 77-81.
- [6] JOVIC A, NIKOLA B. Analysis of ECG records using ECG chaos extractor platform and weka system[C]//Proceedings of ITI 2008 30th Int Conf on Information Technology Interfaces. Cavtat, Croatia: [s.n.], 2008.
- [7] HEINZ-OTTO P, HARTMUT J, DIETMAR S. Chaos and fractals-new frontiers of science[M]. 2nd ed. Berlin: Springer, 2004.
- [8] LU Yue-li, JIANG Di-neng, JIA Xiao-feng, et al. Predict the neurological recovery under hypothermia after cardiac arrest using C_0 complexity measure of EEG signals[C]//30th Annual International IEEE EMBS Conference Vancouver. Canada: IEEE, 2008.
- [9] 李颖洁, 樊飞燕, 朱贻盛. 精神分裂症脑电在不同认知作业下的复杂度分析[J]. *北京生物医学工程*, 2006, 25(5): 470-473.
- LI Ying-jie, FAN Fei-yan, ZHU Yi-sheng. Complexity analysis of schizophrenia EEG during different mental tasks[J]. *Beijing Biomedical Engineering*, 2006, 25(5): 470-473.
- [10] 郁洪强, 赵欣, 李宁, 等. 网络成瘾患者的EEG小波熵与复杂度特征分析[J]. *中国生物医学工程学报*, 2009, 27(2): 157-160.
- YU Hong-qiang, ZHAO Xin, LI Ning, et al. The study on the EEG wavelet entropy and complexity of Internet addiction patients[J]. *Chinese Journal of Biomedical Engineering*, 2009, 27(2): 157-160.
- [11] SABETI M, KATEBI S, BOOSINAI R, et al. Entropy and complexity measures for EEG signal classification of schizophrenic and control participants[J]. *Artif Intell Med*, 2009(3): 3.
- [12] LI Y, TONG S, LIU D, et al. Abnormal EEG complexity in patients with schizophrenia and depression[J]. *Clinical Neurophysiology*, 2008, 119(6): 1232-1241.
- [13] FERENETS R, LIPPING T, ANIER A, et al. Comparison of entropy and complexity measures for the assessment of depth of sedation[J]. *IEEE Trans Biomed Eng*, 2006(53): 1067-1077.
- [14] ABASOLO D, HOMERO R, GOMEZ C, et al. Analysis of EEG background activity in alzheimer's disease patients with Lempel-ziv complexity and central tendency measure[J]. *Med Eng Phys*, 2006(28): 315-322.
- [15] WATANABE T, CELLUCCI C, KOHEGYI E, et al. The algorithmic complexity of multichannel EEGs is sensitive to changes in behavior[J]. *Psychophysiology*, 2003(40): 77-97.
- [16] 尧德中. 脑功能探测的电学理论与方法[M]. 北京: 科学出版社, 2003.
- YAO De-zhong. Electric method and theory of brain function detecting[M]. Beijing: Science Press, 2003.
- [17] XU J, LIU Z, LIU R, et al. Information transmission in human cerebral cortex[J]. *Physica D*, 1997(106): 363-374.

编辑 黄 莘