

符号动力学在心率变异性分析中的参数选择*

宋爱玲 黄晓林[†] 司峻峰 宁新宝

(南京大学电子科学与工程学院, 生物医学电子工程研究所, 近代声学教育部重点实验室, 南京 210093)

(2010年4月30日收到; 2010年5月24日收到修改稿)

时间序列的符号动力学信息熵 H_k 因其计算简单快速, 对数据量要求小, 而被应用于心率变异性 (heart rate variability, HRV) 分析, 然而符号化的参数选择至今却并未形成统一标准. HRV 作为典型的生理信号, 存在着极大的个体间差异和非平稳性, 要获得稳健的一致性分析, 在符号化过程中必须考虑符号化参数 α 与序列本身均值、标准差的综合影响. 文中, 首先以仿真噪声序列为对象, 考察了 3 个参数对于 H_k 的影响及三者相互之间的关联性, 研究表明当满足特定关系时, H_k 的曲线簇收敛于反映序列动力特性的 H_{k-up} ; 随后在对 15 例心跳间隔序列的分析中, 验证了 H_{k-up} 在消除个体间差异及减弱非平稳干扰影响两方面都优于 α 取固定值时的研究结果.

关键词: 符号动力学, 熵, 心率变异性

PACS: 05.45.Tp, 87.19.Hh

1. 引言

健康人的逐拍心率或心跳间隔不是整齐划一的, 而是存在着微小的波动^[1,2], 俗称心率变异性 (heart rate variability, HRV). HRV 的存在绝不仅是外界刺激的结果, 而是由心脏作为复杂的多输入的非线性系统的本质造成, 即使是健康人在安静状态下也普遍存在^[3]. 目前所知的健康人的心率至少受 3 个因素——自动窦性节律、交感神经以及副交感神经的影响^[1-3], 而去神经支配后的自动窦性节律是相对稳定的, 因此, HRV 主要反映了自主神经系统对心脏功能的调控. 以往研究表明: 自主神经系统与心血管疾病以及心源性猝死有着密切联系, 例如由于交感张力的升高与副交感张力的降低, 使室颤阈降低, 易引发室速、室颤和猝死, 而 HRV 的降低则可以作为急性心梗后死亡的强有力的独立的预测因子^[4,5]. 同时, 对于许多其他与自主神经功能相关的非心脏性疾病, 如糖尿病、睡眠呼吸停止、癫痫等, 以及运动医学等方面, HRV 也具有重要的临床诊断或研究价值^[6-8].

HRV 的分析方法主要包括线性方法和非线性方法两大类^[1,2]. 其中如频域分析等传统的线性方法至今仍被广泛应用于临床诊断. 20 世纪 90 年代

以后, 随着对心率波动信号的非线性本质的认识的逐步深入^[9], HRV 的非线性分析成为研究的热点^[2,10]. 尽管 HRV 的非线性分析目前在临床应用中还未形成普遍的统一的标准, 但是作为心脏动力系统非线性本质的刻画, 非线性方法仍然显示出其独特的重要性^[11].

在 HRV 的非线性分析中, 有一类方法——与符号动力学相结合的熵分析方法, 由于其对数据长度要求小, 计算简单快速, 而显得极具吸引力. 在该类方法中, 首先将时间序列, 例如心跳间隔序列, 在幅度域上进行符号化 (粗粒化) 编码, 从而将模拟量的时间序列简化为由有限个符号构成的符号序列, 然后考察某一固定长度下时间序列中各种子串模式的出现概率, 并对其求信息熵. 幅度域的符号化一方面极大地提高了计算速度, 另一方面, 如果方式选择得当, 尽管会丢失一些细节信息, 但却能在保存动力系统的本质特性的同时大大降低噪声的影响, 因此选择恰当的符号化依据成为符号化熵分析方法中的关键. 符号化依据的选择一般有 3 种方式: 1) 以整个时间序列的均值或方差为基准作为符号划分的阈值^[12,13], 或直接定义临界点^[14], 例如 Kurths 等^[12,13] 提出的符号动力学信息熵; 2) 以时间序列的局域波动作为符号划分的依据, 例如李锦等^[15] 提出的基本尺度熵; 3) 以时间序列中相邻两点

* 国家自然科学基金 (批准号: 60701002) 资助的课题.

[†] 通讯联系人. E-mail: xlhuang@nju.edu.cn

的差值(一阶差分)作为符号化的依据,例如卞春华等^[13,16]提出的符号序列熵.后两种方法中,无论是以局域波动作为符号划分的依据,还是以时间序列的一阶差分作为编码的依据,其结果都忽视了时间序列的整体信息或低频特性^[17],例如在对高斯白噪声的基本尺度熵计算中,其熵值远远达不到理论值,而不同类别信号间的差异也因为低频信息的丢失而大大减小,不利于信号的区分.对于这两种方法的具体讨论,可参考文献[13,15—17].本文中主要讨论的是第一种方法,即符号动力学信息熵的参数影响及其选择依据.

2. 符号动力学信息熵

符号动力学信息熵最早由 Kurths 等^[12,13]提出,其计算方法如下:对于数据长度为 N 的时间序列 $\{x_i\} (1 \leq i \leq N)$,将其转换成符号序列 $\{s_i\} (1 \leq i \leq N)$,其中 $s_i \in \{0, 1, 2, 3\}$,具体的转换方法如下:

$$s_i(x_i) = \begin{cases} 0 & \mu < x_i \leq (1 + \alpha)\mu \\ 1 & (1 + \alpha)\mu < x_i < \infty \\ 2 & (1 - \alpha)\mu < x_i \leq \mu \\ 3 & x_i \leq (1 - \alpha)\mu \end{cases}, \quad (1 \leq i \leq N), \quad (1)$$

式中 μ 代表时间序列 $\{x_i\}$ 的均值, α 则为一个特殊参数.

对于符号序列 $\{s_i\}$,其中所有长度为 m 的子串的集合可表示为 $\{W_j\} (1 \leq j \leq N - m + 1)$, $W_j = (s_j, s_{j+1}, \dots, s_{j+m-1})$,考察 $\{W_j\}$ 的分布特性,即对出自符号表 $\{0, 1, 2, 3\}$ 的长度为 m 的所有可能的 4^m 个符号串,统计其在 $\{W_j\}$ 中出现的次数 $C(l) (1 \leq l \leq 4^m)$,并估计出概率

$$p(l) = \frac{C(l)}{N - m + 1}, \quad (2)$$

对其计算信息熵

$$H_k = - \sum_{p(l) > 0} p(l) \log p(l), \quad (3)$$

从而得到时间序列 $\{x_i\}$ 的符号动力学信息熵. H_k 反映了时间序列在符号化以后各种长度为 m 的子串的模式丰富程度和分布特性,出现的符号串模式越多、分布越分散,则熵值越高;符号串模式越少、分布越集中,则熵值越低.例如,对于时间上无相关的完全随机序列,如高斯白噪声序列,其长度为 m 的子串在字符表 $\{0, 1, 2, 3\}$ 中取值的可能组合共有 4^m 种,且各种组合模式出现的概率基本一致,因而计

算得到的熵值达到最大值 $\log(4^m)$,而对于有特定时序结构的序列,其子串的模式则会远远小于 4^m ,同时在分布上也很可能出现不均衡现象,从而熵值小于 $\log(4^m)$.

关于计算 H_k 时参数的选择,子串的长度 m 的增加会导致计算量及必要数据长度的指数增加,而对于结果却无本质影响,因此,一般取到 3 就足够了^[12,13];时间序列长度 N 的选择,为使得 H_k 具有统计学意义,一般要求 $N \gg 4^m$ ^[12];而符号化过程中的参数选择则非常重要.从(1)式可以看出,序列 $\{x_i\}$ 的均值 μ 和参数 α 都会影响符号化的结果,另外,尽管(1)式中没有直接反映,但是可以推断,序列 $\{x_i\}$ 的标准差(standard deviation, SD)也会影响符号化的结果.事实上,这三个参数并非完全独立,而是相互关联的.在以往研究中, Kurths 和 Wessel 等只给出了在研究 HRV 时, α 可以取 0.1^[12]或者取 0.03 到 0.07 之间^[13],然而,HRV 作为典型的生理信号,个体间差异大,且同一个体也往往存在着极大的非平稳性,固定的参数选取会导致分析结果受到心跳间隔序列的均值、标准差等参数的影响,而无法获得稳健的一致性结果.因此,在下面的章节中,将进行详细的符号化参数分析.

3. 基于仿真噪声序列的符号化参数分析

本节将以仿真噪声序列为例详细讨论 3 个参数的影响以及相互之间的关联性.

3.1. μ, α 和 SD 对 H_k 的影响

首先以高斯白噪声序列为对象,考察 μ, α 和 SD 对于 H_k 计算的影响.为简化起见,事先将长度 $N = 1000$ 的噪声序列进行了零均值和归一化标准差的预处理,然后人为地改变其 μ 或者 SD,并计算不同 α 取值时的 H_k ,计算中,取 $m = 3$.结果如图 1 所示,其中图 1(a)为固定 $SD = 1$ 时, H_k 随 μ 和 α 变化的曲线,图 1(b)为固定 $\mu = 10$ 时, H_k 随 SD 和 α 变化的曲线.

从图 1 中可见, H_k 的值随着时间序列的 μ, SD 以及参数 α 的变化而变化,即不同的序列均值、标准差和参数 α 都会影响计算出来的 H_k ;另一方面,对于多数 α - H_k 曲线,都存在着一个极大值,不同曲线的极大值基本相等,即收敛于一个固定值,这里

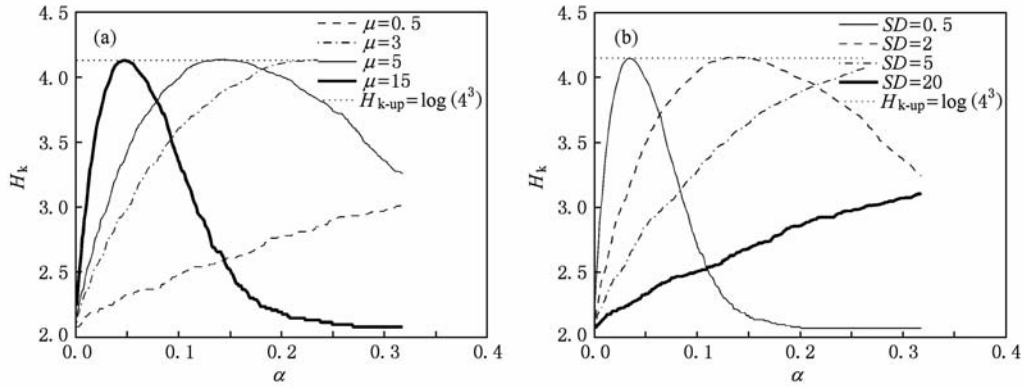


图1 高斯白噪声序列 H_k 随 α, μ, SD 的变化情况 (a) 固定 SD 时, 随 μ 变化的 H_k - α 曲线簇, (b) 固定 μ 时, 随 SD 变化的 H_k - α 曲线簇

定义为 H_{k-up} , 该值非常接近理论推导值 $\log(4^3)$, 这说明: 在 μ, SD, α 三个参数的特定组合下, H_k 能达到理论值, 从而能够正确的反映时间序列的内在动力学特性; 而参数取值不当时, H_k 则可能会远远小于 $\log(4^3)$, 从而不能真实地反映序列的复杂程度.

除了高斯序列, 我们还对 $1/f$ 噪声和布朗噪声做了相同的计算. 计算结果显示, 对于这两种噪声序列, 其 H_k 随着 μ, SD 以及 α 的变化情况与图1类

似, 并且 α - H_k 曲线簇也存在着收敛的极大值 H_{k-up} —— $1/f$ 噪声为 3.5, 布朗噪声为 1.9. $1/f$ 噪声和布朗噪声都存在着时间相关性, 两者的区别在于尺度因子的不同. 对于存在时间相关的序列, 其长度为 m 的子串会集中在相对有限的模式, 因而其 H_k 理论上应小于无时间相关性的高斯序列, 且尺度因子越大, 相关性越强, 其 H_k 则越小, 这与我们的计算结果是相符的.

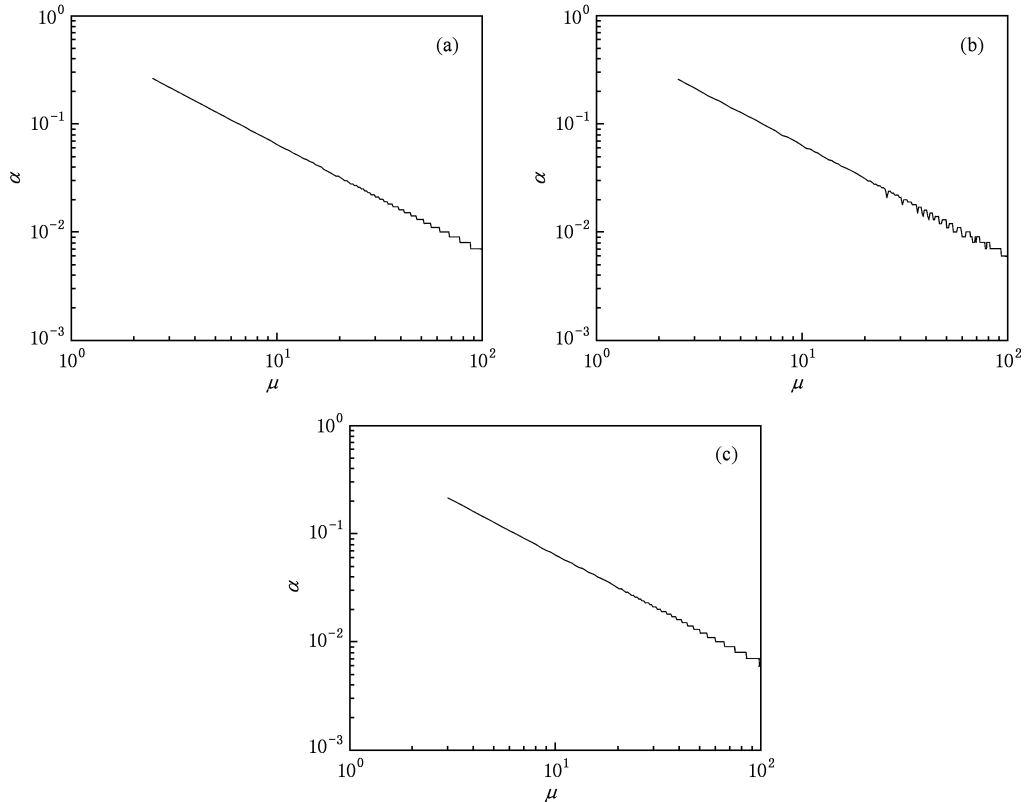


图2 获得 H_{k-up} 时的 μ - α 关系曲线 (a) 高斯白噪声, (b) $1/f$ 噪声, (c) 布朗噪声

3.2. 最佳参数设置

既然对于三种典型噪声序列,其 α - H_k 曲线簇的极大值都收敛于 H_{k-up} ,并且通过高斯序列已经验证了 H_{k-up} 与理论值非常接近,我们推断 H_{k-up} 可以作为时间序列复杂性的一种稳健的反映.下面仍然以上述三种仿真噪声序列为对象,考察要获得 H_{k-up} , μ , SD 以及 α 之间必须满足的关系.简单起见,这里将 SD 固定为 1,着重研究 μ - α 关系.固定 μ 时,SD- α 关系可以类似方法获得,在此不赘述.

图 2 中分别是 2000 点的标准差为 1 的高斯序列(a), $1/f$ 噪声序列(b), 布朗噪声序列(c) 获得 H_{k-up} 时, μ - α 关系双对数曲线.从图 2 中可见,三种序列与 H_{k-up} 对应的 μ - α 关系双对数曲线都呈现出非常明显的线性,因此,针对三种噪声各 150 例独立的序列,进行了 μ - α 关系双对数曲线的线性回归并作了 F 检验,详细结果列于表 1.

表 1 基于仿真噪声序列的 μ - α 双对数曲线线性回归

序列	回归斜率 (均值 \pm SD)	回归截距 (均值 \pm SD)	线性回归 的 F 检验
高斯噪声	-1.000 \pm 0.006	-0.392 \pm 0.036	$P < 0.01$
$1/f$ 噪声	-1.001 \pm 0.013	-0.409 \pm 0.088	$P < 0.01$
布朗噪声	-0.998 \pm 0.032	-0.373 \pm 0.315	$P < 0.01$

从表 1 中可看出, μ 在 2.5—100 之间时,经 F 检验证实,三种噪声 μ - α 关系的双对数曲线的线性关系显著,且线性拟合的斜率均近似为 -1,截距也都在 -0.4 附近.三种噪声之间,无论是拟合的斜率还是截距,差异均不显著(t 检验 $P > 0.05$),这说明:对于这三种存在不同时序结构的序列,要获得反映其动力学复杂性的 H_{k-up} ,当 $SD = 1$ 时,其 μ - α 关系遵循基本同等的规律,即满足

$$\log \alpha = -\log \mu - 0.4. \quad (4)$$

由(4)式可以看成在计算符号动力学信息熵 H_k 时参数选取的最佳方案.有了这个标准,按(1)式符号化时,时间序列单纯由均值、标准差的线性变化带来的影响基本被消除.

4. HRV 的分析应用

在本节中我们研究的序列来自于 Chaos 在 2008

年发起的专题讨论“正常心率是混沌的吗?”所提供的数据库*,其中包括 5 例正常窦性心率(normal sinus rhythm, NSR)(年龄 30 ± 10),5 例充盈性心衰(congestive heart failure, CHF)患者(年龄 59 ± 9),5 例房颤(atrial fibrillation, AF)患者(年龄不详)的长时(20—24 h)心跳间隔序列.生理信号的个体差异和非平稳性,一直是阻碍其分析获得稳健的一致性结果的一个重要原因.例如在 HRV 研究中,采集心电图数据时不同个体之间所处状态的差异,以及单个个体在持续心电采集时心跳间隔序列的非平稳性,都会严重影响到分析结果.本节中,针对上述两种情况,分别研究了利用最佳参数设置后,符号动力学信息熵对于 HRV 分析的改善.

首先对所有 15 例数据的最初 2000 点,分别利用固定参数,取 $\alpha = 0.05$ (仿照文献[13]), $\alpha = 0.01$,以及第 3 节得到的最佳参数设置方法,计算了各自的符号动力学信息熵,分别记为 H_{k1} , H_{k2} 和 H_{k-up} ,结果如图 3 所示.图 3 中,横轴是 15 例数据的编号,其中 1—5 为 NSR(n1nn—n5nn),6—10 对应 CHF 患者(c1nn—c5nn),11—15 则对应 AF 患者(a1nn—a5nn);纵轴是符号动力学信息熵 H_k 的计算值;图中‘○’代表应用最佳参数方案获得的熵值 H_{k-up} ,而‘*’和‘+’则分别代表不考虑个体间均值差异而固定取 $\alpha = 0.05$ 时的熵值 H_{k1} 和 $\alpha = 0.01$ 时的熵值 H_{k2} .

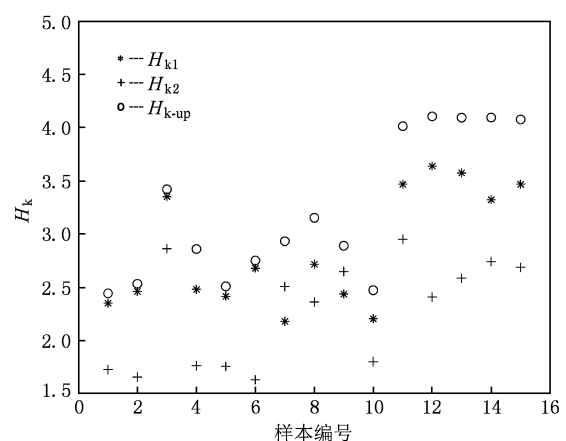


图 3 心跳间隔序列的 H_{k1} , H_{k2} 和 H_{k-up} 结果比较

由图 3 可见,三组样本的 H_{k-up} 相对于 H_{k2} 都有明显的改变,而对于 ID 为 4 以及 7—15 的序列, H_{k-up} 与 H_{k1} 也有显著的差别,特别是,采用 H_{k-up} 时,

* <http://www.physionet.org/challenge/chaos/>

AF 患者完全与 NSR 和 CHF 区分开来,同时 AF 类别内的个体差异也明显减小.为进一步说明利用 H_{k-up} 更能反映序列的动力学本质而尽可能排除均值等线性因素的影响,图 4 中分别给出了 CHF 组(a)和 AF 组(b)的 $\mu-H_k$ 的散点图,图中实线代表拟合

直线.从图 4 中可见, H_{k-up} 与序列均值的线性相关性相对于 H_{k1} , H_{k2} 减小,这说明,采用标准化参数后,组内成员之间尽管采样时存在均值、标准差等状态的不同, H_{k-up} 仍然能真实地反映动力系统的复杂性,从而获得一致性的结果.

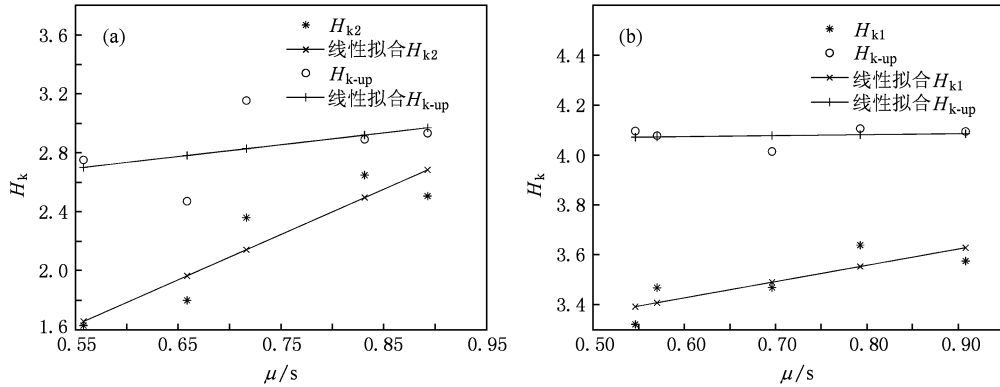


图 4 心跳间隔序列的 $\mu-H_k$ 散点图 (a) CHF 组, (b) AF 组

接下来,则通过对长时数据的分段分析来考察在同一个体中, H_{k-up} 减弱均值波动影响的能力.对于每例数据,我们都将其按 2000 点一段分为 31 段,分别计算每段的均值、 H_{k1} 、 H_{k-up} ,再考察每例样本 H_{k1} , H_{k-up} 的波动范围及其与均值之间的相关性,对于线性相关显著 ($P < 0.05$) 的则对其进行线性拟合,比较线性拟合的斜率,图 5 画出了具有代表性的 a1nn 数据的分段分析结果,其中图 5(a) 为熵值随分段时间的波动,其横轴代表分段的序号,纵轴代表符号动力学信息熵,图 5(b) 则为其 $\mu-H_k$ 的散点图,实线代表线性拟合.由图 5 可见,就 a1nn 而言, H_{k-up} 的波动范围明显小于 H_{k1} ,且 H_{k-up} 对于 μ 的线性依赖性也显著降低.三组共 15 例样本的全部结果列于表 2 中,其中, Δ 代表参数的值域, k 代表线性拟合的斜

率, NS 代表线性相关不显著 ($P > 0.05$).

从表 2 中可以看到,就熵值的变化范围而言,除了 n3nn, c1nn, c4nn (表中有下划线的数字) 以外,其他样本中 H_{k-up} 波动都是小于 H_{k1} 的;而从 $\mu-H_k$ 的相关性分析及线性拟合斜率来看,除了 n3nn (表中有下划线数字) 的 H_{k-up} 随 μ 的变化比 H_{k1} 大, n4nn, n5nn, c1nn, c4nn 两种熵值与 μ 均无显著线性相关,其他样本都表现出 H_{k-up} 随 μ 的变化减小或者与 μ 的线性相关性的削弱.这两个方面都说明,采用最佳参数设置后的 H_{k-up} 对于 HRV 的复杂性描述更稳定,更不易受均值波动的影响.而在三组样本中,以 AF 组的改善效果为最好,这是由于 α 取固定值 0.05,即是文献[13]在完全不考虑 AF 病人的前提下针对正常范围的平均心率和标准差情况挑选过

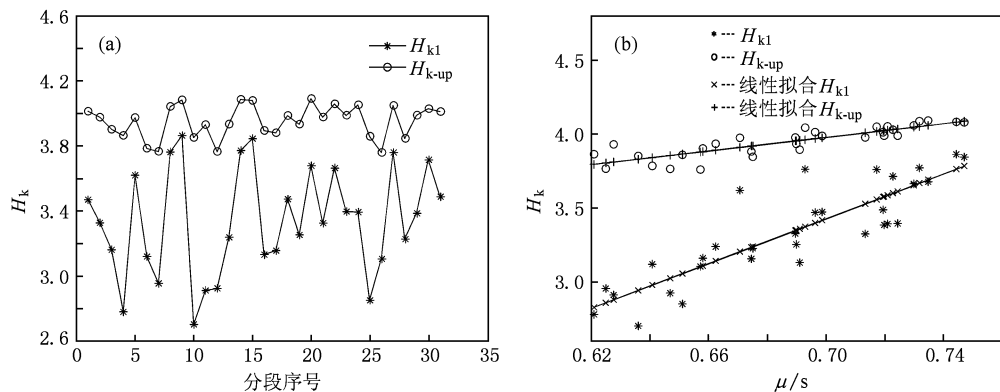


图 5 a1nn 的长时心跳间隔序列分段分析 (a) H_k 随分段时间的波动情况, (b) $\mu-H_k$ 散点图

的,所以就 NSR, CHF 而言该 α 与最佳参数偏离不是特别大,而 AF 组其均值和标准差较 NSR 和 CHF 组有较大偏离,因而 α 取 0.05 的选择严重偏离其最佳参数,从而造成看起来 H_{k-up} 对于 AF 组的改善效果最好.

表 2 长时心跳间隔序列分段分析的 H_k 波动范围及与 μ 的线性相关性分析

数据	$ \Delta_{H_{k1}} - \Delta_{H_{k-up}} $	$k_{\mu-H_{k1}}$	$k_{\mu-H_{k-up}}$
n1nn	0.208	2.125	1.986
n2nn	0.530	1.876	1.693
n3nn	-0.114	2.346	2.579
n4nn	0.694	NS	NS
n5nn	0.074	NS	NS
c1nn	-0.690	NS	NS
c2nn	0.018	-3.308	2.648
c3nn	0.165	-4.558	NS
c4nn	-0.363	NS	NS
c5nn	0.165	2.391	1.511
a1nn	0.831	7.594	2.317
a2nn	0.278	1.257	NS
a3nn	0.185	-0.256	NS
a4nn	0.323	0.517	NS
a5nn	0.345	1.643	NS

图 6 为三种噪声和三组 HRV 的 H_{k-up} 值的比较.

由图 6 可见, AF 组可以完全与 NSR 和 CHF 组区分开来,而且与高斯噪声接近,而 NSR 和 CHF 组的 H_{k-up} 则处于 $1/f$ 噪声和布朗噪声之间. 这说明, AF 患者的心跳间隔序列存在着严重的时间相关性缺失,其随机性与白噪声接近,而 NSR 和 CHF 患者的心跳间隔序列则存在着时间相关性,这种相关性比 $1/f$ 噪声略强,而比布朗噪声又要弱,这与以往通过

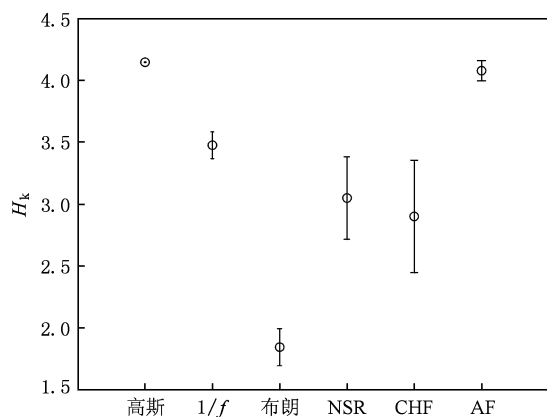


图 6 仿真噪声和 HRV 的 H_{k-up} 对照

功率谱或去趋势波动 (detrended fluctuation analysis, DFA) 等方法得到的研究结果是一致的^[18,19].

5. 讨论和结论

以往研究表明,心率波动序列作为确定性系统产生的非随机序列,仅由概率分布函数来描述是不够的,其排列的先后顺序也蕴含了大量的内在动力学特性^[1],符号动力学信息熵则能从一定程度上衡量这种时间前后的关联性,从而评价时间序列的复杂性.符号动力学信息熵计算简单快速,对数据量要求不大,但是符号化参数的选取一直以来并未形成统一标准,而 HRV 作为生理信号的一种,具有个体间差异大和非平稳等特点,因此,如何尽量消除个体差异使参数具有可比性,以及如何消除非平稳干扰的影响而获得一个稳健的一致性的分析结果,使得为符号化建立一个参数选取的最佳方案成为必须解决的问题.

本文对符号动力学信息熵 H_k 的符号化方法进行了参数分析:首先通过三种代表性的仿真噪声序列(高斯噪声、 $1/f$ 噪声、布朗噪声),研究了符号化参数 α 与序列本身均值、标准差的影响及相互关联性,结果表明当 α 与均值、标准差满足特定关系((4)式)时, H_k 会收敛于反映真实动力学特性的 H_{k-up} ;然后,针对三组共 15 例心跳间隔序列验证了利用最佳参数设置方案的 H_{k-up} 无论是消除不同个体间状态差异,还是减弱同一个体在长时心电采集中的基础心率的影响,都优于固定 α 的 H_k . 最后,通过比较三组 HRV 与三种噪声的 H_{k-up} ,认为 AF 组心跳间隔序列在时间相关性上与高斯白噪声基本接近,而 NSR 组和 CHF 组则更接近 $1/f$ 噪声,即 NSR 和 CHF 组的心跳间隔序列存在着类似 $1/f$ 噪声的时间相关性,这与以往研究认为正常人 HRV 包含 $1/f$ 成分^[18,19]的观点是相符的.

同时我们也看到,在 NSR 和 CHF 组的 H_{k-up} 分段分析结果中, H_{k-up} 仍然存在着不可忽略的波动,既然基础心率等线性非平稳干扰带来的影响已经消除,那么 H_{k-up} 的波动很可能是由心脏动力系统本身固有的低频特性所引起,2000 点的心跳间隔序列不足以包含这种低频成分,这很可能就是导致 NSR 组和 CHF 组之间仅由 H_{k-up} 无法区分的重要原因. 要详细研究这些低频成分,需要有大量的长时心跳间隔序列,并且要在采集过程中严格标

定测试者各个时间段的状态,这将是我們进一步工作的方向.

感谢南京邮电大学地理和生物信息学院马千里博士对文中噪声的获得和处理提出的宝贵建议.

-
- [1] Camm A J, Malik M, Bigger J T, Breithardt G, Cerutti S, Cohen R J, Coumel P, Fallen E L, Kennedy H L, Kleiger R E, Lombardi F, Malliani A, Moss A J, Rottman J N, Schmidt G, Schwartz P J, Singer D H 1996 *Eur. Heart J.* **17** 354
- [2] Kleiger R E, Stein P K, Bigger J T 2005 *A. N. E.* **10** 88
- [3] Ivanov P C, Chen Z, Hu K, Stanley H E 2004 *Physica A* **344** 685
- [4] Kleiger R E, Miller J P, Bigger J T, Moss A J 1987 *Am. J. Cardiol.* **59** 256
- [5] Malik M, Farrell T, Camm A J 1990 *Am. J. Cardiol.* **66** 1049
- [6] Huikuri H V, Makikallio T H, Airaksinen K E J, Seppanen T, Puukka P, Raiha I J, Sourander L B 1998 *Circulation* **97** 2031
- [7] Stein P K, Kleiger R E 1999 *Annu. Rev. Med.* **50** 249
- [8] Zhuang J J, Ning X B, Zou M, Sun B, Yang X 2008 *Acta Phys. Sin.* **57** 2805 (in Chinese) [庄建军、宁新宝、邹鸣、孙飙、杨希 2008 物理学报 **57** 2805]
- [9] Li C, Tang D K, Fang Y, Sun J T, Ding G H, Poon C S, Wu G Q 2009 *Acta Phys. Sin.* **58** 1348 (in Chinese) [李程、汤大侃、方勇、孙锦涛、丁光宏、Poon C S、吴国强 2009 物理学报 **58** 1348]
- [10] Ning X B, Bian C H, Wang J, Chen Y 2006 *Chin. Sci. Bull.* **51** 385
- [11] Wessel N, Riedl M, Kurths J 2009 *Chaos* **19** 028508
- [12] Kurths J, Voss A, Saparin P, Witt A, Kleiner H J, Wessel N 1995 *Chaos* **5** 88
- [13] Wessel N, Ziehmann C, Kurths J, Meyerfeldt U, Schirdewan A, Voss A 2000 *Phys. Rev. E* **61** 733
- [14] Liu X F, Yu W L 2008 *Acta Phys. Sin.* **57** 2587 (in Chinese) [刘小峰、俞文莉 2008 物理学报 **57** 2587]
- [15] Li J, Ning X B, Wu W, Ma X F 2005 *Chin. Phys.* **14** 2428
- [16] Bian C H, Ma Q L, Si J F, Wu X H, Shao J, Ning X B, Wang D J 2009 *Chin. Sci. Bull.* **54** 4610
- [17] Huang X L, Cui S Z, Ning X B, Bian C H 2009 *Acta Phys. Sin.* **58** 8160 (in Chinese) [黄晓林、崔胜忠、宁新宝、卞春华 2009 物理学报 **58** 8160]
- [18] Kobayashi M, Musha T 1982 *IEEE Trans. Biomed. Eng.* **29** 456
- [19] Peng C K, Havlin S, Stanley H E, Goldberger A L 1995 *Chaos* **5** 82

Optimum parameters setting in symbolic dynamics of heart rate variability analysis*

Song Ai-Ling Huang Xiao-Lin[†] Si Jun-Feng Ning Xin-Bao

(Key Laboratory of Modern Acoustics of Ministry of Education, Institute of Biomedical Electronic Engineering,
School of Electronic Science and Engineering, Nanjing University, Nanjing 210093, China)

(Received 30 April 2010; revised manuscript received 24 May 2010)

Abstract

The Shannon entropy H_k in symbolic dynamics analysis of time series has less data demand and can be complemented easily, and therefore is applied to heart rate variability (HRV) analysis. However, the criterion has not been established as to the parameter α , which is used during the transformation into symbols. Like all other physiological signals, great differences between individuals as well as nonstationarity are usual in HRV. Therefore, to achieve a robust and consistent analysis, the parameter α should be considered together with the mean and standard error of the series. In this paper, the integrative effects on H_k of the three parameters were studied firstly in simulation time series, and it was suggested that under certain conditions, the clusters of H_k approach a convergent upper boundary named H_{k-up} , which reflects the intrinsic dynamics of the time series. Then, in the heart beat interval series analysis of 15 subjects, the abilities of H_{k-up} to alleviate the impacts brought by both difference between individuals and nonstationarity were testified.

Keywords: symbolic dynamics, entropy, heart rate variability

PACS: 05.45.Tp, 87.19.Hh

* Project supported by the National Natural Science Foundation of China (Grant No. 60701002).

[†] Corresponding author. E-mail: xlhuang@nju.edu.cn