

摘要

生物电信号的分析对于评估人体机能的状态具有重要的意义。本文提出了两种新的生物电信号的分析方法，期望能够更有效的对生物电信号的不可逆性及耦合程度进行分析，并应用于实践。本文的工作主要有以下三个方面：

一、青老年脑电信号时间不可逆的分析及对比

本文提出了一种新的使用过程的前向概率和后向概率的方法来计算符号相对熵，并利用符号相对熵来估计熵增的方法，该方法可以证明脑电信号具有时间不可逆特性。最后应用该方法对青老年脑电信号进行数值计算及对比，结果是老年人的平均能量损耗显著高于年轻人，证明符号相对熵可以作为一个物理过程不可逆程度的度量参数，这对评估脑功能具有重要的指针作用。

二、癫痫异常脑电信号不可逆性分析

本文应用基于符号化相对熵算法对癫痫异常脑电信号及正常脑电信号进行数值计算及对比，结果是癫痫患者的平均能量损耗显著高于正常人的，证明符号相对熵可以作为衡量脑电信号是否异常的参数。

三、多参量生物电信号的耦合分析

本文提出了一种新的时间序列耦合信息分析方法—符号化部分互信息熵。应用该算法对生物电信号计算并进行假设检验，结果表明多参量的生物电信号各参量间具有耦合关系，清醒期的生物电信号耦合程度显著高于睡眠期的。证明符号化部分互信息熵可以用来分析时间序列间的耦合信息，而且生物电信号的耦合程度可以作为深入理解人体各器官相互作用的重要参考。

关键词：脑电信号，符号化相对熵，多参量生物电信号，符号化部分互信息

Abstract

In order to accurately predict and diagnose health of body organs, effective way to collect and analyze the biological electrical signals is needed. This paper put forward two new biological electrical signal analysis methods for getting more effective analysis on the irreversibility and coupling degree of biological electrical signals. What's more, it was applied in practice. This thesis was mainly focused on the following three aspects:

Firstly, time irreversibility of electroencephalogram signals were analyzed.

In this thesis, we showed that the trajectories of electroencephalogram (EEG) possesses the character of time reversal asymmetry, which could provide information about the entropy production of EEG. We developed a new method to estimate symbolic relative entropy whose entropy production using forward and backward trajectories. Then we used this method to dispose and analyze the EEG of younger and elder subjects. It was turned out this method working and the average energy dissipation could be used as a parameter to detect nonequilibrium.

Secondly, abnormal epilepsy electroencephalogram signals were analyzed based on symbolic relative entropy.

Symbolic relative entropy algorithm was used in the analysis of abnormal epilepsy electroencephalogram. It showed that abnormal epilepsy electroencephalogram had different average energy dissipation from normal EEG's. Therefore, the average energy dissipation could be used as a parameter to judge if the EEG signal was normal.

Thirdly, the coupling of multivariate bioelectricity signals was analyzed.

Symbolic partial mutual information was proposed in this paper, which was based on partial mutual information. This algorithm could be used to analyze the coupling between multivariate time series. We used this method to dispose and analyze the sleeping MBS (multivariate bioelectricity signal) and waking one. it could be reached that the coupling of waking MBS was obviously larger than sleeping one's. Finally statistical hypothesis testing was done to prove this method working and the average energy dissipation could be used as a parameter to detect nonequilibrium.

Keywords: EEG, symbol-relative entropy, multivariate bioelectricity signal, coupling analysis

目录

第一章 绪论	1
1.1 生物电信号概述	1
1.2 生物电信号研究现状	3
1.3 论文研究内容及结构安排	4
第二章 生物电信号基础理论	6
2.1 生物电活动基本理论	6
2.1.1 脑电活动的生物学基础及信号采集	6
2.1.2 心电活动的生物学基础及信号采集	10
2.1.3 肌电活动的生物学基础及信号采集	13
2.1.4 生物电信号的采集	15
2.2 生物电采集中干扰信号及降噪处理	16
2.2.1 生物电采集中干扰信号	16
2.2.2 生物电信号降噪处理方法	16
2.3 符号动力学概述	18
2.3.1 符号动力学基本理论	18
2.3.2 符号化映射方法及符号信息熵	19
2.4 常用生物电数据库	20
2.5 小结	22
第三章 基于符号化相对熵的脑电信号时间不可逆性分析	23
3.1 熵增原理与时间不可逆性	23
3.2 基于符号相对熵脑电信号时间不可逆性研究	24
3.2.1 相对熵理论	24
3.2.2 基于符号序列的相对熵方法	25
3.3 数据分析	27
3.3.1EEG 实验数据	27
3.3.2 实验方法与结果分析	27
3.4 小结	35
第四章 基于符号化相对熵的癫痫异常脑电信号分析	36
4.1 癫痫异常脑电信号病理机制及研究现状	36
4.1.1 癫痫异常脑类疾病病理机制及研究意义	36
4.1.2 癫痫异常脑电信号研究现状	42
4.2 异常脑电信号不可逆性分析及意义	44
4.2.1 基于符号化相对熵的异常脑电信号不可逆性分析基本理论	44
4.2.2 算法流程	46
4.3 异常脑电信号不可逆性分析实例仿真	47
4.3.1 实验数据	47
4.3.2 实验方法	47
4.3.3 实验结果与分析	47
4.4 小结	55
第五章 基于符号化部分互信息的多参量生物电信号分析	57
5.1 信息熵研究现状	57
5.1.1 互信息基本原理	57
5.1.2 部分互信息基本原理	60
5.2 时间序列耦合分析	61

5.2.1 基于符号化部分互信息算法的时间序列耦合分析基本理论.....	61
5.2.2 算法流程	61
5.3 多参量生物电信号耦合分析实例仿真	62
5.3.1 实验数据	62
5.3.2 实验方法	62
5.3.3 实验结果及分析	63
5.4 小结	71
第六章 总结与展望	72
参考文献	74
附录 1 攻读硕士学位期间撰写的论文	78
附录 2 攻读硕士期间参加项目	79
致谢	80

第一章 绪论

1.1 生物电信号概述

在生命活动过程中，生物的器官、组织和细胞等会发生电位和极性的变化，这是正常生理活动的表现。生物电信号^[1]定义为在生命活动过程中，在生物体内或体表产生的各种电位和电流。心电、脑电、肌电、动作电位、细胞膜电位、眼电信号、诱发电位信号等等都属于生物电信号。

2000 多年前人类就发现存在生物电现象^[2]，并加以利用。18 世纪末，L.伽伐尼首次提出“动物电”的观点。20 世纪初，W.艾因特霍芬首次直接测量到微弱生物电信号的存在。1922 年，H.S.加瑟和 J.埃夫兰格首先使用示波器研究生物的动作电位，开创了现代电生理学的先河。1939 年，A.L.霍奇金和 A.F.赫胥黎使用微电极测量出乌贼大神经纤维膜内外的电位差。从此电生理学理论进入迅速发展阶段。近年来，生物电信号的测量和采集技术已经趋于成熟。

生物体内的生物电现象是一种正常的生理活动，一定的生理活动对应于一定的生物电反应，即生物电在统计意义上存在一定的规律性。因此，从统计学角度来分析生物电，可以为我们提供一定的生理活动的信息，据此可以判断一个生理机制是否处于健康有序的状态。目前生物电的研究已经广泛应用于临床医学，如脑电图、心电图、肌电图等等，医务工作人员可以根据这些信息来诊断病情。在可预见的未来，生物电在仿生学、信息控制、物联网等领域也会产生巨大的作用。

大脑^[3]作为人体最重要的器官之一，其健康状况对人类的影响巨大，生活中的大脑一词指的是端脑。端脑是中枢神经系统的最高级部分，由 140 亿个脑细胞构成，据估计人体每天约有 10 万个脑细胞死亡，而且越不用脑死亡越多。随着经济的不断发展，人们的工作强度和生活压力也在不断增加，提高人类大脑的学习和休息效率意味着提高人们的工作效率，人类能够在相同的时间内创造更多的社会价值和经济效益。因此越来越多的人们开始关注大脑的健康以及如何提高大脑的工作效率。脑电信号的采集、特征提取和分析是近年来医学领域研究的重点。脑电信号分析在信息技术及仿生学方面的重要意义也不可忽视，目前主要有三个研究方向：脑电波理疗程序和配套仪器的开发；脑电波多媒体技术的开发；基于脑电波的传感器和“天网”的构造。天网是指以脑电波为基础的超媒体通信网络，它由脑电波网络、有线电话网络、无线通信网络、计算机网络、卫星网络组成的大通信网络，使得人机器的直接交互成为可能。未来脑电波信息技术不仅可以为人类提供理疗保健功能，还可以传递包括

温度、触觉、嗅觉、痛觉等感觉信息，为人类提供更全面真实的效果。因此脑电信号分析和研究具有极其重要的科研价值、社会价值、经济价值，带来新的技术革命，能够极大的推动社会和人类文明的进步。

近年来心血管疾病的发病率越来越高，这是因为随着经济的发展，人们生活质量和饮食条件不断提高，但是又不能合理科学安排膳食及运动，现在人群普遍存在身体超重、蛋白质缺乏、膳食纤维缺乏、盐量过度的问题。合理膳食及适度运动对心血管疾病的防治是非常重要的。心脏对人体的重要性不言而喻，它的健康状况将会直接影响到人的身体健康质量。从以往的统计材料来看，心脏疾病一直是危害人类健康的主要疾病之一，并且导致死亡的几率很高^[4,5]。目前常见的血管类病重包括冠心病、高血压、运动猝死、心绞痛及高血脂等。现在常见的一个认识误区是心血管疾病的发病集中于老年人群体，但是调查表明，这些疾病已经以越来越高的比例困扰着所有年龄群体。在这种情况下，对心脏的功能状况进行定期的检测与评价以及对心脏疾病进行及时的预报和有效的诊断对人类健康就显得非常重要。及时有效的对心血管疾病进行检测和预防可以有效的降低发病率。心电信号^[6]是一种典型的生物电信号，具有振幅、相位、频率等特征要素，能够准去的反映心脏的健康状况，因此广泛的应用于临床医学，对于心脏类及血管类基本的诊断和治疗有重要意义。目前成人心血管疾病的诊断和治疗已趋于完善，但是对婴幼儿的心脏疾病的诊断和治疗却不尽如人意。人们在关注自身健康的同时，也越来越关注下一代的健康，尤其是优生优育的重要性。据有关部门统计资料显示，先天性心脏病是新生儿及儿童期的导致死亡的主要原因之一。在我国，新生婴儿中患有先天性心脏病的比例约为新生婴儿总数的 0.8%~1.2%之间，这意味着每年我国将会出生约 10 万的先天性心脏病患儿。因此有必要采取措施对胎儿形成阶段的生理和病理情况的做全面深入和细致了解，能给家庭和医生提供有关详尽的胎儿发育的基本信息，以便对妊娠过程做出有效的决策。这主要是因为胎儿形成阶段所具有的质量将直接影响婴儿的健康状况，并对后续的生长发育有深远的影响。因而在围产期^{*}对胎儿和孕妇进行监护受到人们较为广泛的重视^[7]。

由于神经肌肉系统的研究工作对残疾人的康复工程有重要意义，因此生物医学工程领域具有广泛的应用前景。世界各国越来越重视神经肌肉系统的研究工作，并投入大量人力物力。目前最主要的研究手段是针电极肌电信号的分解。肌电信号的分解在临床神经肌肉系统疾病的诊断和治疗领域也有重要应用。临床神经系统疾病常采用神经肌电图检查法，可以对个神经组织的功能状态进行判断和评定。随着研究的深入，肌电信号分析技术越来越成熟，促进

* 围产期是指怀孕 28 周到产后一周这一分娩前后的重要时期。

了各类神经肌肉系统功能障碍的诊断和康复治疗。

1.2 生物电信号研究现状

目前常用的生物电分析方法^[8]主要有以下几类：

一、时域分析方法

时域分析方法是起源最早的生物电分析方法，其基本原理是从时域中直接提取信号特征，具有直观性强、算法成熟等优势，是目前最成熟的生物电分析方法。在生物电信号的定量和定性分析中，特别是在瞬态波形的研究中，包括棘慢波（反映癫痫信息）和梭形波等，生物电信号的某些重要信息在时域上的反映更为突出，所以至今仍广泛应用于实践工作中。目前较为成熟的时域分析方法主要包括：相关分析方法、直方图分析方法、波形识别、过零点分析方法、峰值检测方法及相干平均等。目前临床医学等领域广泛采用了这些时域分析方法，为患者的健康状况的分析和诊断提供依据。

二、频域分析方法

生物电信号的频域分析方法已经日趋完善，现在广泛应用于心电及脑电信号健康状况分析及研究。频域分析的主要手段是功率谱分析，这种方法是从频域提取出信号特征。需要对采集到的时域生物电信号进行时频变换^[8]，把幅度随时间变化的生物电信号变换为生物电功率随频率变化的谱图，从而提取并分析生物电信号的频域特征。经典谱估计和现代谱估计是早期最常用的功率谱分析方法。这两种方法有很大的局限性，没有良好的方差特性，即估计值沿频率轴的起伏剧烈。近年来兴起的现代谱估计技术很好的克服了这种缺点，应用最为广泛的是参数模型法。参数模型法特别适合应用于生物电信号的动态分析，具有高频率分辨率及谱图平滑的优点，有利于参数的自动提取和定量分析。

三、时频分析方法

单纯使用时域分析方法或者频域分析方法都不能很好的对其特征进行提取和分析，这是由于生物电信号本身的特征决定的。因为生物电信号具有时变、非平稳特征。而谱分析要求信号是平稳信号，且对于瞬态信号没有良好的分析作用。生物电信号在不同时刻具有不同频率成分，而时域分析又不能分析信号的频域特征。为了对信号进行更全面的分析，必须对生物电信号的时间和频率结合起来进行处理分析。时频分析方法是把时域和频域结合起来表示信号。目前最为常用的时频域方法方法是小波分析方法。小波变换在高频时使用短窗口，在低频时使用宽窗口，相当于一组中心频率可变、带宽相等的带通滤波器，实现了信号的实时处理。小波分析方法特别适合脑电信号的分析，这是由于脑电信号具有很强的非平稳性。在

不同尺度上，脑电信号中的棘慢波、棘波具有不同的表现，根据这一特点，使用小波变换中的多尺度分析方法能够检测出这些异常波。因此小波变换在生物电信号分析领域具有重要研究意义。另外常用的方法还有多维统计分析方法、非线性分析方法等等。

虽然采用上述的时频域分析方法研究生物电信号取得了很多有重要价值的结论，但是上述方法存在两个局限：第一是对数据的需求量过大^[9]，第二是算法效率不高，不能有效的满足临床诊断的快速精确的要求。随着研究的不断深入，出现了将符号动力学^[10]与熵运算结合的研究生物电的方法。其主要原理是首先是对生物电电信号进行符号粗粒化，然后再对符号化处理得到的信号进行不同的熵运算，达到对生物电信号的特征的有效提取和分析，据此可以研究生物电信号的不可逆性和耦合程度。符号化的主要作用是去除噪声的影响，同时能够保持原始序列的大尺度信息。比较典型的改进熵算法有基本尺度熵^[11]、样本熵^[12]和多尺度熵^[13]等等。

本文研究主要目的是如何能够更有效的分析脑电信号的不可逆性以及多参数生物电信号的耦合程度，主要方向有：一是对脑电信号不可逆程度的分析研究，例如在相对熵基础之上进行改进，提出了符号化相对熵算法；二是对多参数生物电信号的耦合程度进行分析研究，提出了符号化部分互信息算法；两种算法对生物电信号的不可逆性和耦合程度有很好的特征分析，对数据的需求量要求不高，并且算法具有高精度和高效率，能够更好的应用于临床的诊断。

1.3 论文研究内容及结构安排

本文在前人的工作基础之上，主要对脑电信号时间不可逆性和多参量生物电信号耦合程度进行了研究，所做的主要工作有：

(一) 针对传统算法的不足并且根据脑电信号的特点，提出了改进的符号化相对熵算法，该算法可以对生物电信号的不可逆性进行分析，指出了青老年的脑电信号的不可逆性存在差异，并给出了青老年符号相对熵值的变换范围，能够为脑电活动生理过程算法处于积极有序状态的判断提供依据，未来可应用于生命状态分析、临床医学、生物电传感器等领域。详见第三章。

(二) 本文第四章应用基于符号化相对熵算法对癫痫异常脑电信号及正常脑电信号进行数值计算及对比，结果是癫痫患者的平均能量损耗显著高于正常人的，证明符号相对熵可以作为衡量脑电信号是否异常的参数。

(三) 针对传统互信息分析方法的缺陷以及算法的复杂性，提出了新的生物电信号耦合程

度分析方法—符号化部分互信息。主要方法是首先对原始序列进行符号化，对符号化的序列提取部分互信息。针对 matlab 处理循环效率低下的不足，算法的概率统计部分采用了 C 和 matlab 编程结合的方式。该算法未来可用于人体工作和休息等状态的改进。详见第五章。

根据上文阐述，全文共分为六章内容。第一章介绍了生物电信号的基础知识、研究现状、本课题所作的工作以及本论文的结构安排，第二章首先介绍了多种生物电信号生物学基础，然后介绍了生物电信号的采集方法以及在采集过程中存在的干扰类型，并给出了相应的降噪预处理的方法。最后简单介绍了目前比较常用的标准心电信号数据库。第三章首先对传统的相对熵算法以及脑电信号不可逆研究现状做了介绍，在此基础上提出了符号相对熵的计算方法，对青老年脑电信号的时间不可逆性的分析和比较，得到了有意义的结论。第四章应用基于符号化相对熵算法，对癫痫异常脑电信号及正常脑电信号进行数值计算及对比，结果是癫痫患者的平均能量损耗显著高于正常人的，证明符号相对熵可以作为衡量脑电信号是否处于健康或平衡状态的参数。第五章对传统的耦合信息分析方法做了介绍，并提出了符号化部分互信息熵算法。该算法能够有效的对三种及以上相互耦合的序列进行耦合分析，并使用该方法对脑电、心电、肌电多参量生物电信号进行了分析，证明可以有效的获得耦合信息。第六章对本文的研究工作做出总结，并指出了本文存在不足以及需要下一步改善的地方。

第二章 生物电信号基础理论

2.1 生物电活动基本理论

2.1.1 脑电活动的生物学基础^[14]及信号采集^[15]

人类大脑是世界上最灵敏最精密的器官。端脑由一百四十亿个脑细胞构成，蕴含了无数待开发的资源。端脑分为左、右半脑两部分。研究表明，左脑和右脑各有分工。人类左脑主要控制知识、判断、思考等显意识功能。人类右脑具有自主性，主要是根据接收到的讯息衍生出创造性的讯息。人类日常的学习等行为主要集中于对左脑的训练，而对右脑相关的功能未给予足够的重视，一般人右脑的五感都受到左脑理性的控制与压抑。科学已经证明，大脑皮层是一个整体的导体，其作用是传导神经元活动产生的微弱的电位变化到大脑表面，EEG 记录大脑表面的电位信息，体现了大脑的生理病理信息。所以在大脑的表面安放电极就可以记录下来大脑在工作时产生的 EEG。因此通过对人体无创伤的 EEG 来获取脑部健康状况信息。本节我们先研究大脑的结构及功能、EEG 产生机制、EEG 导联放置方法以及 EEG 的基本波。

人类大脑的构成复杂，主要包括脑干、间脑、小脑和端脑四部分，结构图如图 2.1 所示：

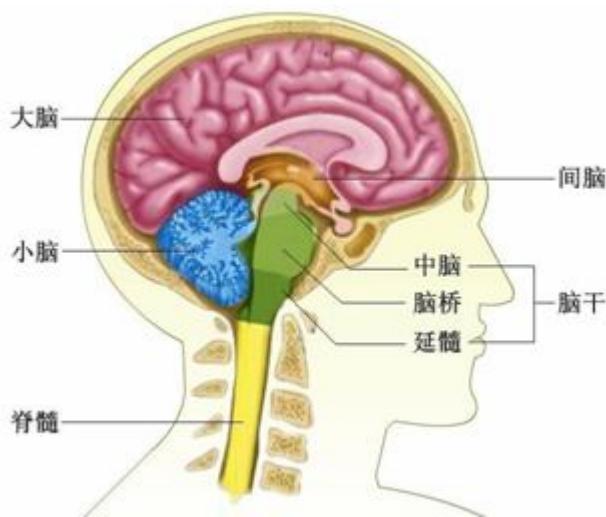


图 2.1 大脑结构

脑干：主要功能是维持生命体征，例如心跳、呼吸等基本生理功能。

间脑：主要负责管理血压以及内脏等活动。

小脑：主要作用是控制机体运动、协调身体平衡。

端脑：是实现脑的高级功能的神经中枢，其健康状况影响着人的智力，是语言、情感等生命活动的控制中心。

神经元细胞是大脑的基本单位。神经元细胞由三部分构成：细胞体、轴突、树突，其基本结构示意图如图 2.2 所示：

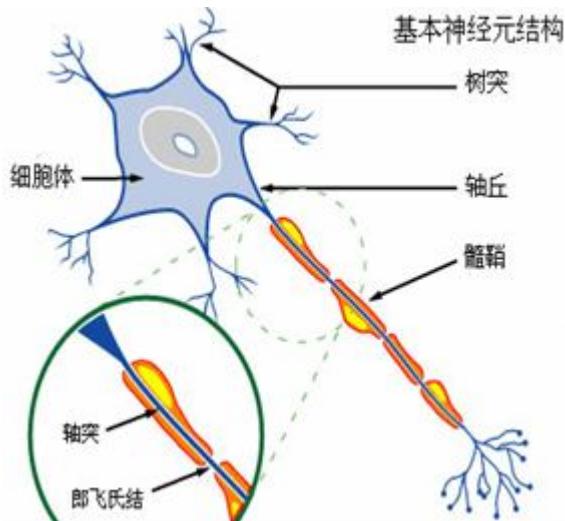


图 2.2 基本神经元结构示意图

细胞体：主要由细胞核、细胞膜、细胞质、线粒体、内质网和核糖体构成，是神经元的基础结构。

轴突：神经元的索状突出部分，其主要作用是传导电位和冲动到其他神经元，属于传出纤维。

树突或者神经末梢：神经元的树枝状细小突出，主要作用是与其他细胞相连，接受刺激，实现神经元间的感知和传递，是传入纤维。

脑电信号产生的生物机制如下：在大脑的生理活动过程中，大脑皮层的神经元会进行传导电位活动。神经元细胞膜内外离子浓度发生变化会产生电位活动。在普通状态下，细胞外的钠离子浓度较高，使得细胞电位是内负外正。在受到刺激时，细胞膜的通透性变强，大量钠离子进入到细胞内部，导致细胞内电位上升，从而变为内正外负。继而细胞关闭纳通道，钾离子流出细胞，导致膜外电位迅速回升，回到内负外正的状态。这一系列的电位变化就产生出神经冲动并通过神经纤维传导出去。大量神经元传导信号混合在一起，通过在头皮上设置电极就可以检测到脑电信号 EEG，EEG 是头皮上两个电极之间电位差值的波形图。EEG 反映了脑组织活动状态，是神经系统疾病的诊治的客观指标。

EEG 信号采集导联放置如图 2.3 所示：

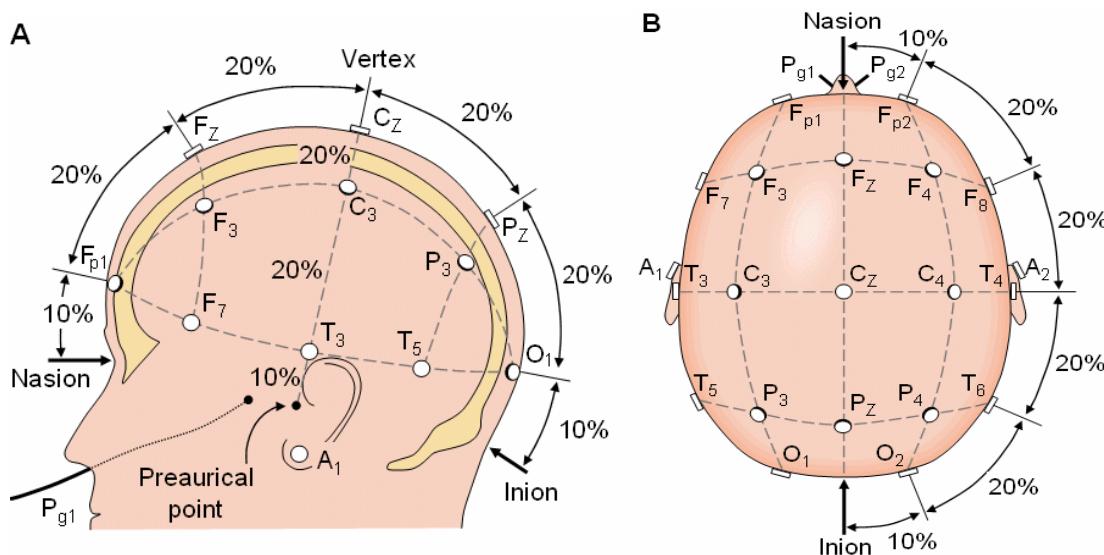


图 2.3 国际 10-20 电极系统

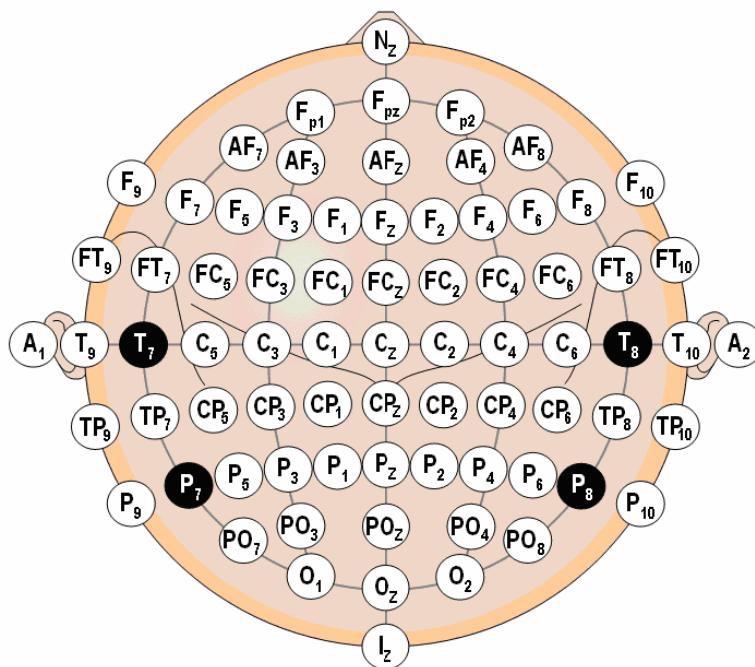


图 2.4 美国 EEG 研究会关于标准系统电极位置和命名

(其中 T7、T8、P7、P8 例外)

目前,临幊上采用的电极安放法为国际 EEG 图学会所建议的 10-20 系统标准电极安放法,这种电极安放方法和大脑物理构造相符合并且各重要区域都有电极分布。从鼻根(眼睛下缘)到脑后枕骨外凸点中线,依次放置电极 F_p , F_z , C_z , P_z , O 。电极名称依次为: 额极、额点、中央点、顶点、枕点, T 为颞点, A 为耳垂电极, Z 为中线电极。 F_p 在鼻根到后枕骨的 10%

处, F_Z , C_Z , P_Z 间距分别为 20%, O 在后枕骨到鼻根的 10% 处, 从左耳前点到右耳前中线上一次放置 $T3$, $T4$, $C3$, $C4$ 。依次为: 左颞中、右颞中、左中央、右中央。 $T3$ 在左耳前点到右耳前点的 10% 处, $T4$ 在右耳前点到左耳前点的 10% 处, $T3$, $T4$, Cz , $C3$, $C4$ 间距分别为 20%。因此该电极安放系统叫做 10-20 系统。电极下标序号左半球为奇数右半球为偶数。

根据频率、波幅和电位的不同, EEG 可划分为半弧状波、锯齿状波、正弦波、类正弦波、双峰波、棘波、尖慢波等^[16]。临幊上常采用频率作为划分参数, 频率是单位时间内出现的周期数, 用 Hz 表示, 周期指的是相邻的波峰到波峰的时间或波谷到波谷的时间。临幊上按频率分为: β (beta)波、 α (alpha)波、 θ (theta)波、 δ (delta)波, 其波形如图 2.5 所示:

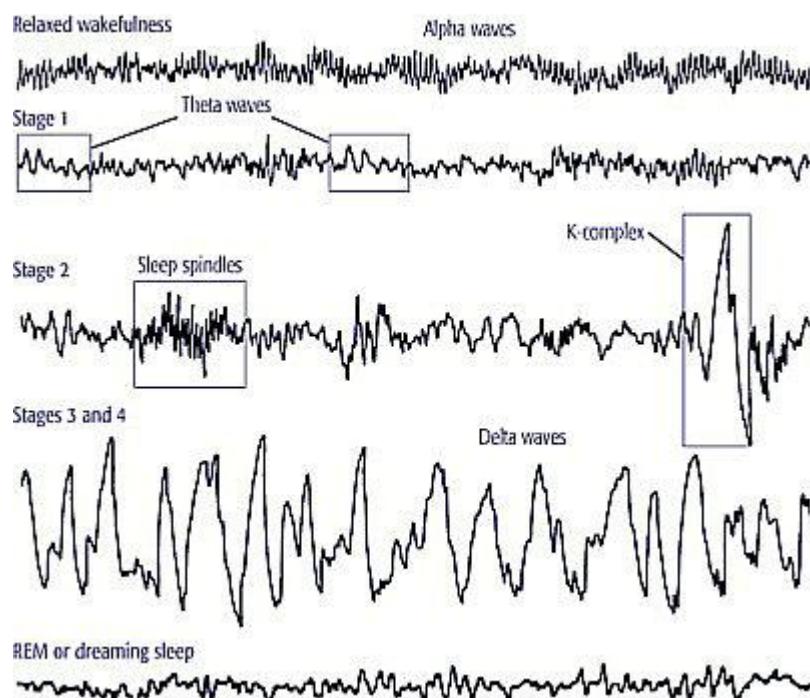


图 2.5 正常人的 EEG

EEG 的四种基本波的频段和波幅特性如表 2.1 所示。

表 2.1 EEG 的四种基本波

名称	波幅	频段
β 波	δ 波的一半	13Hz-30Hz
α 波	25 μ V-75 μ V	8Hz-13Hz
δ 波	20 μ V 以下	0.5Hz-4Hz
θ 波	20 μ V-40 μ V	4Hz-7Hz

β 波 频率为 13~30Hz，波幅为 5~20 μ V，较 α 波频率高、幅度低，故称之为快波。在额、颞和中央区较为明显，在皮层张力增高即神经元兴奋性增高导致的情况下出现，比如在情绪波动较大、突然受到刺激时急剧增多。

α 波 是正常成年人的基本节律，清醒安静闭眼时出现，在顶、枕区最为明显。当睁眼注意思考或者接受其他刺激时^[16]α 波消失，出现低幅快波，称为 α 波阻断。正常人头颅两侧对应导联得到的 α 波频率一般相同。

θ 波 慢波，在顶区、额颞区活动较为明显，疲劳时会出现，轻睡时，θ 波首先会出现在两侧额部，逐渐消失。皮质下的病变可导致产生爆发性 θ 波，属不正常表现。正常情况下出现的百分比小于 15%。

δ 波 慢波，在婴儿至儿童期或者成年人的正常睡眠期出现在额区，代表皮层张力降低，在婴儿直至少年期属正常。病灶性 δ 波意味着皮层发生了病变，如果出现在双侧和爆发性 δ 节律，则提示通常是皮层下发生了病变。

小波变换特别适用于脑电信号的分析，α 波、β 波、θ 波、δ 波在小波的不同尺度上表现不一。实验证明，在受试者睁眼和闭眼时，会有 α 阻断现象，这一现象在尺度 $2^1 \sim 2^3$ 上出现明显变化；癫痫患者发病时期脑电信号常见为 3c/s 棘慢波形式，在尺度 $2^1 \sim 2^5$ 会有显著改变。影响 EEG 的因素有很多，年龄和精神活动状态均会对 EEG 造成影响。普遍来讲，相当于中老年群体，青少年对外界刺激反应较为敏感，脑电波容易出现异常，但是恢复也比较快。EEG 信号的采集存在很多的干扰因素，比如眼球运动、呼吸、血管搏动、哭泣、肌肉收缩等生理过程都可能对 EEG 信号的检测造成干扰。因此如何克服各种干扰因素的影响也是 EEG 研究领域中的重要课题。

2.1.2 心电活动的生物学基础及信号采集

人类心脏的外形类似于桃子，其大小与成人拳头相当，呈倒置圆锥体状。从外形来看心脏可分为前面、侧面和后面。心脏由左心房、左心室、右心房和右心室组成。分隔心房和心室的标志是冠状沟。冠状沟位于心脏的底部，横着绕心脏一周。心脏作为人体血液循环的动力器官，是人体机械活动强度最大的人体器官，其相应的生物电信号强度也最大。心脏主要由主要由心肌组织组成。心肌组织是心脏生物电活动的基本单位，自律性、收缩性、兴奋性、传导性是心肌组织四大特性。心脏生物电活动具有三大特点：自律性、兴奋性和传导性。自律性是指心肌细胞能够在缺乏外来刺激的情况下自发的产生除极等生物电活动。兴奋性是指在外来或内在的刺激作用下，心肌细胞能够产生一次兴奋即一个电位变化，并向邻近的组织

传导。传导性是指心肌细胞将兴奋即动作电位传递给其他心肌细胞。

一、心肌细胞的生物电活动机制

每个心肌细胞都被一层质膜包围，将细胞内外环境相互隔开。在代谢产生的过程中，心肌细胞质膜两侧的离子浓度环境会产生变化。主要表现是心肌细胞内的 Na^+ 离子的浓度低，而 K^+ 离子浓度高；即离子分布处于不平衡状态。这种不平衡状态的存在，使心肌细胞膜两侧产生了静息电位，让细胞内外两侧产生了大小约为-90mV的电位差，处于内负外正状态。在不平衡的状态下的刺激下，随着膜内外电位差的改变，细胞质膜的通透性产生周期性的变化。细胞膜首先变成导通状态，对 Na^+ 离子通透性的增加；由于扩散和电位梯度的共同作用，造成大量 Na^+ 离子向细胞内流动，进而形成 Na^+ 离子流，使得该细胞和其相邻细胞间产生一个电位差变化。这种电位差变化会刺激周围相邻心肌细胞进行传导兴奋。因此，兴奋一旦形成就可以在心肌细胞间不停地进行传导。由于周期性的作用，当细胞膜回归到静息状态时，由于离子分布不平衡状态形成的心肌细胞动作电位即告结束，重新处于静息状态。心肌细胞电活动中的离子穿膜电位的生理过程如图 2.6 所示：

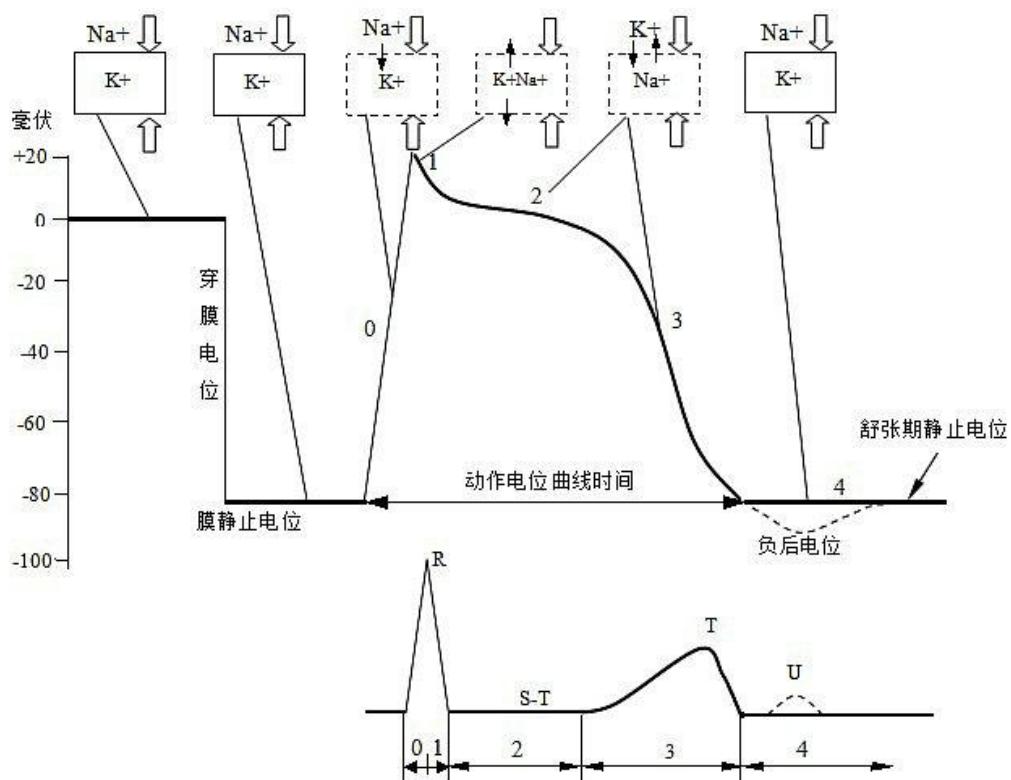


图 2.6 上行代表单细胞的穿膜动作，中行代表心肌纤维的动作电位曲线，下行是与之对应的心电图

二、心电图生理学

在心脏正常的情况下，心肌动作电位开始于窦房结。窦房结是人体心脏的原发性起搏兴

奋点，它位于右心房和上腔经脉的交界处。窦房结有节律的产生兴奋，平静时大约为 60~100 次/min。然后兴奋在传导组织中进行传播，最终将兴奋传至心室外膜，从而完成一次完整的兴奋生成和传导过程。心脏各部位在兴奋过程中的电变化都有一定规律，这种规律性在传播分析、时间和途径方面都有体现，具有一定周期性，体表心电信号随着心脏搏动周期也完成一个周期。人体体表两个部位之间电位差并随时间而变化，心电图是使用心电描记器记录这种电位变化而得到的图形，即为 ECG (Electrocardiograph)^[17]。心电图自 1993 年引入临床医学，心电图在心脏类疾病的临床诊断有重要价值。随着人们生活质量的提高，社会对心血管类疾病的诊断和治疗也越来越重视，在心电图分析领域投入大量的人力物力，心电信号的采集、分析与诊断技术均日渐成熟。心电图分析十分广泛的应用于现代医学领域，创造了巨大的社会价值和经济价值。

健康条件下，正常心脏在一次完整的除极过程中产生的 ECG 信号通常包括 P 波、QRS 波以及 T 波^[18]，如图 2.7(a)所示；图 2.7 (b)为 ECG 各个组成部分的波形。

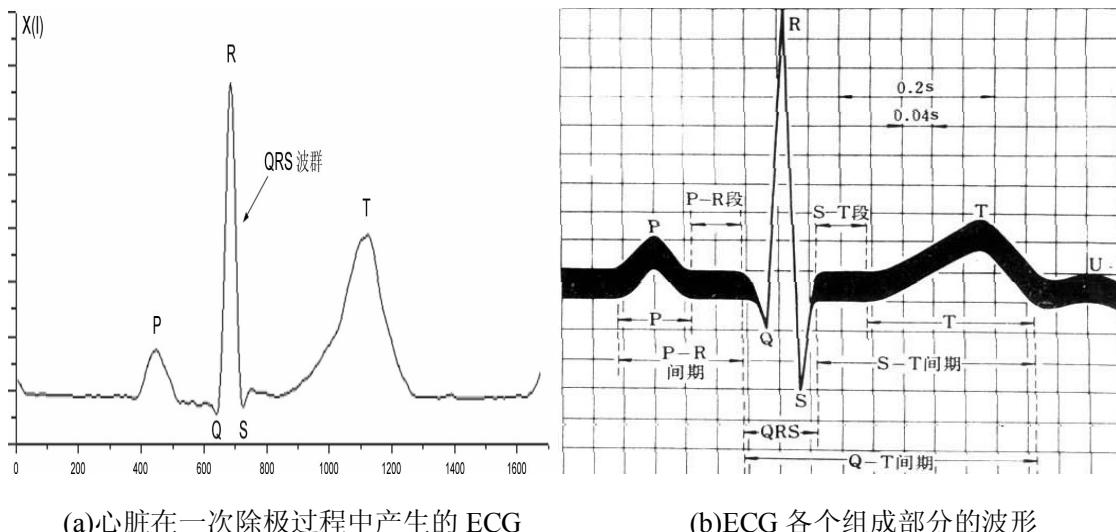


图 2.7 正常人的典型心电图

三、心电信号的采集

人类从 19 世纪中叶开始对 ECG 信号进行采集，人类首次心电图记录是 Waller 用毛细管静电计记录下人体心电图波群。生理学家 Einthoven 研制出心电图机，为心电图研究作出了开拓性的贡献。从 20 世纪初起，心电检测系统经历了由“弦线型放大器”到“电子管”到“晶体管”到“集成块”的发展，逐渐发展成为精巧实用的心电检测设备。从 20 世纪 80 年代开始，多导心电同步检测设备开始应用于临床医学，这是心电检测技术的又一里程碑。

国际标准十二导联系统具有兼容标准导联和单体导联的优点，是目前临幊上采集 ECG 广

泛采用的方法。该系统具有 12 个导联，分别记为 I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6，能够比较全面和细致的记录心脏电活动。健康年轻人的十二导联 ECG 如图 2.8 所示。



图 2.8 一名健康青年人的同步十二导联 ECG

图 2.8 中 aVR、aVL 和 aVF 导联为 Goldberger 改进设计的加压肢体导联；I, II, III 导联为 Einthoven 设计使用的标准肢体导联；其他的六个导联主要是放置在心脏胸前壁的体表上。导联 V2 和 V5 在临床诊断和研究中常作为主要诊断参考依据。

2.1.3 肌电活动的生物学基础及信号采集

肌电信号的研究已经成为康复工程各种技术实现的关键，体表肌电信号临床医学常用生物电只有，蕴含了很多肌肉运动信息，它由中枢神经系统发出信号，并在神经纤维的传输过程中不断加强，最终在体表能够检测和采集到有效的肌电信号。

肌电信号^[19]来自于运动神经元。运动神经元是中枢神经的一部分，连接着多条肌纤维，这些部分合在一起，构成运动单位。其示意图如图 2.9 所示：

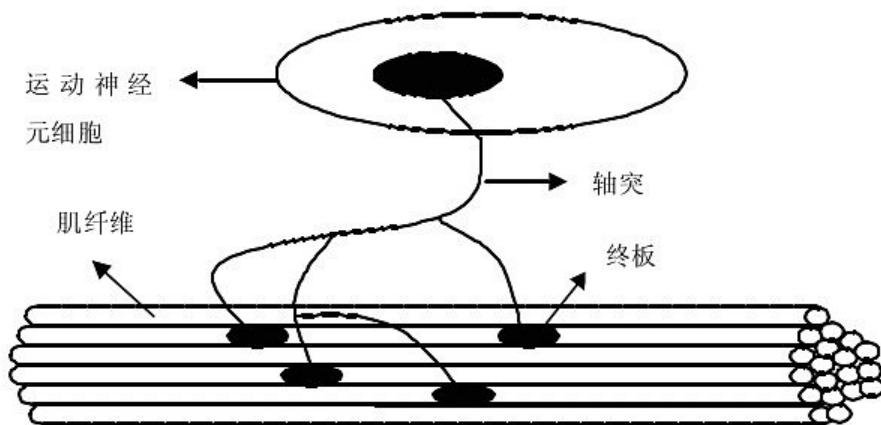


图 2.9 运动单位

一、肌电信号生物学基础

运动单位能够被随意的激活，不同运动单位的动作电位组成肌电信号。肌电信号产生是生物机制如下：肌电—骨骼肌兴奋时^[20]，首先中枢神经系统发出驱使肌肉收缩的信息—运动电位，由于肌纤维动作电位的产生、传导和扩布，而发生电位变化称为肌电。用适当的方法将骨骼肌兴奋时发生的电位变化引导、记录所得到的图形，称为肌电图(electromyogram, EMG)。肌电信号产生机制如图 2.10 所示：

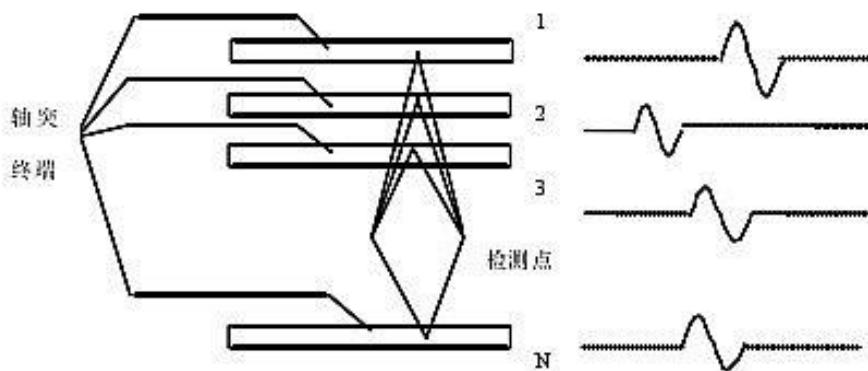


图 2.10 肌电信号产生机制

二、肌电信号的采集^[20]

1907 年 Piper 首次记录到人类臂肌肉的电位差。20 世纪中叶以来，随着神经肌肉生理学研究的发展和电子计算机的创新，人们在肌电信号的检测和记录技术也日趋完善，能够对神经肌肉功能和肌电信号进行更为深入细致的研究。经过半个多世纪的发展，国内外肌电信号采集系统已经能够较好的解决原始肌电信号的提取放大及分析工作，肌电信号采集系统原理图如图 2.11 所示：

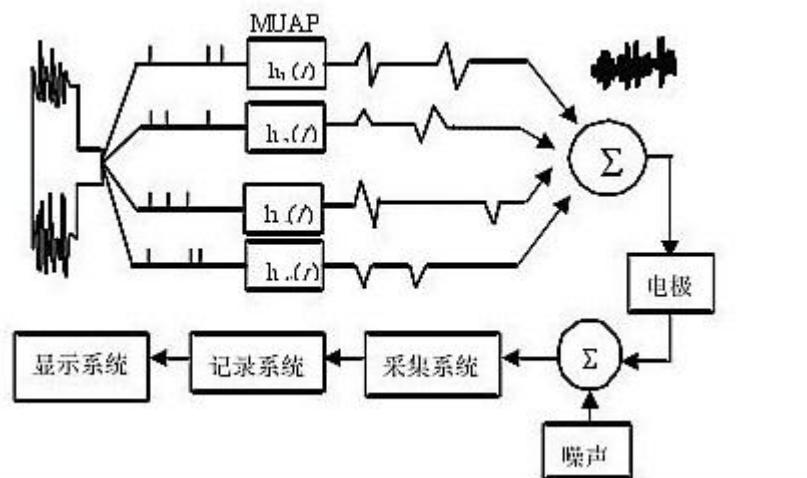


图 2.11 肌电信号采集系统

2.1.4 生物电信号的采集^[21]

生物电信号及其微弱，多在毫伏级或微伏级，所以极易受环境的影响。生物电信号具有噪声强信号弱的特点，基线漂移、工频干扰、不同生理信号间的信号影响等都属于常见的这背景噪声，均具有较强的干扰性，因此生物电信号容易淹没在各种干扰引起的强背景噪声之中。在灵敏度、分辨率、共模抑制、抗干扰等方面对生物电信号采集电路有更高的要求。因此生物电信号采集系统应该包括差分信号放大、滤波处理、以及其他信号调理电路。常用的生物电信号采集系统一般包括图 2.12 所示的 10 个部分：

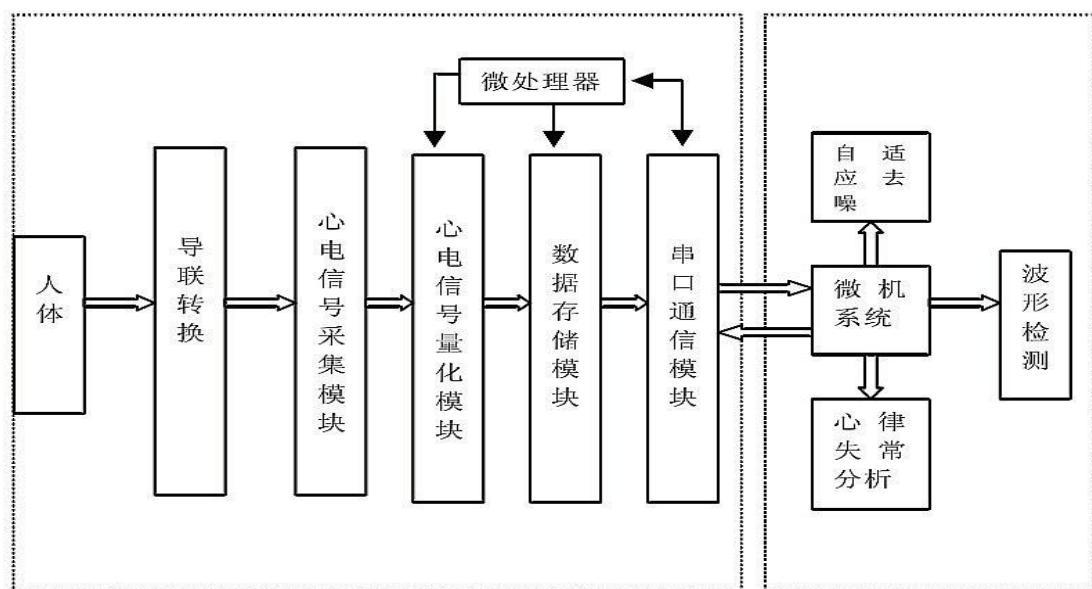


图 2.12 生物电信号采集系统框图

2.2 生物电采集中干扰信号及降噪处理

2.2.1 生物电采集中干扰信号

从生物电信号的形成和采集中可以看出，采集到的原始生物电信号比较微弱，表现在幅度一般保持在十微伏到五毫伏之间；能量分布在零点零五到一百赫兹之间，其中主要的能量是集中在 0.5~40Hz 这一范围。因此在采集过程中，采集得到的心电信号会混入噪声。这会给下一步的分析和诊断带来很大的影响。所以在进行分析和诊断之前，就必须对采集到的原始生物电信号进行必要的预处理，降低噪声的影响^[22]。在实际情况中，干扰生物电信号的噪声主要有：工频干扰、基线漂移、肌电干扰、电磁设备干扰和电极接触干扰等。使用软件滤波算法来提高系统的信噪比也是必要的抑制干扰的手段^[22]，这是由于使用模拟滤波器进行滤波并不能把干扰信号全部削掉。因此在送入上位机之前，首先会对采样数据使用模拟滤波器滤波，某些干扰信号会出现在模拟滤波器之后，对于之后的干扰就需要采用滤波算法来消除^[22]。自适应滤波基于较好的适应性，相比与目前比较流行的小波处理方法，自适应滤波运算量小且收敛速度快^[23]。

2.2.2 生物电信号降噪处理方法

下面介绍工频干扰和基线漂移的特点以及相应的自适应滤波处理方法：

二、自适应滤波抑制工频干扰^[24]

来自外界的各种干扰中，以工频干扰对信号的影响最大，甚至可能会淹没微弱的生物电信号。工频干扰噪声主要是由基波频率为 50Hz 及其各谐波组成的，它是由于市电以电磁波形式向外辐射而产生。因此必须采取有效的手段来抑制 50Hz 工频。工频干扰下的原始 ECG 信号如图 2.13 所示：

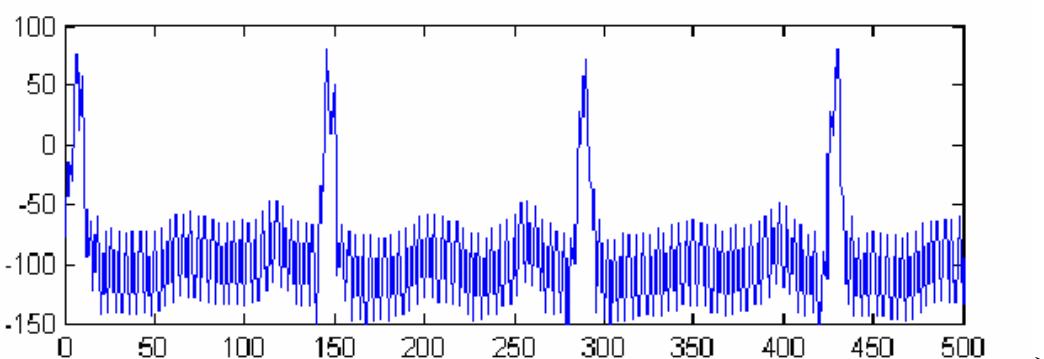


图 2.13 工频干扰下的原始 ECG 信号

在实际的降噪处理过程中，传统采用的方法主要有：50Hz 陷波、自适应滤波^[24]等。对工频干扰下的原始ECG信号采用自适应滤波算法进行处理，采用二阶滤波器，采样频率为200Hz ($\mu = 0.1, \gamma = 1$)。滤波后信噪比提高了22db，滤波后的结果如图2-14所示：

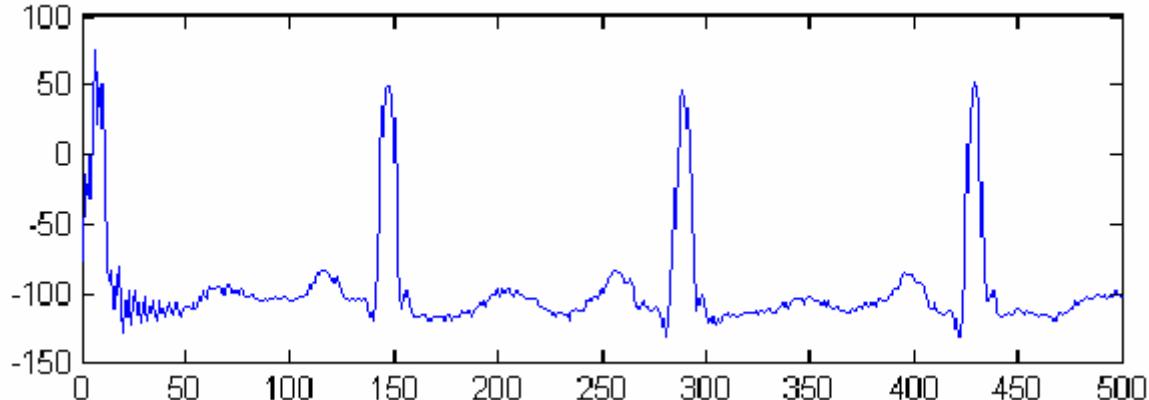


图 2.14 自适应滤波后的 ECG 信号

(2) 自适应滤波抑制基线漂移^[25]

基线偏移是生物电干扰信号中最难消除的，基线漂移的频谱主要是0.05Hz~2Hz之间，属于低频干扰。基线漂移的来源主要有：转接盒接触不良、电极板的分极化效应、电极与皮肤接触不佳、外来行为的干扰，被采集者肢体的移动、呼吸以及其他运动对导联影响。临床数据表明，引起基线漂移的主要因素之一是呼吸波^[23]，因此抑制呼吸波是抑制基线漂移有效手段。图2.15是比较典型的由呼吸引起ECG的基线漂移。

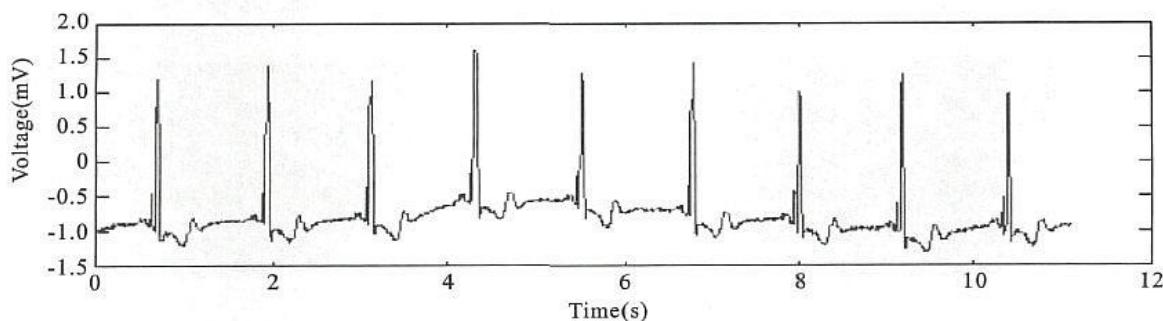


图 2.15 由于呼吸引起 ECG 基线漂移

由上图可知，由于呼吸导致体表不稳定引起ECG的基线漂移非常明显。为了尽量的降低基线漂移对ECG波形的影响，在临床中采用的应对措施有：一方面使用较好的电极材料，使电极牢固地与皮肤吸附，被采集对象尽量避免大呼吸运动^[26]。另一方面是对其进行适当滤波处理。消除ECG信号中的呼吸波干扰需要经过以下几个处理过程：首先对ECG信号与呼吸波信号进行预处理^[23]，再使用自适应对消器分别对它们分别进行处理，收敛后输出的信号分

量是近于完全匹配的，即已经消除呼吸波干扰。自适应滤波后的 ECG 信号如图 2.16 所示：

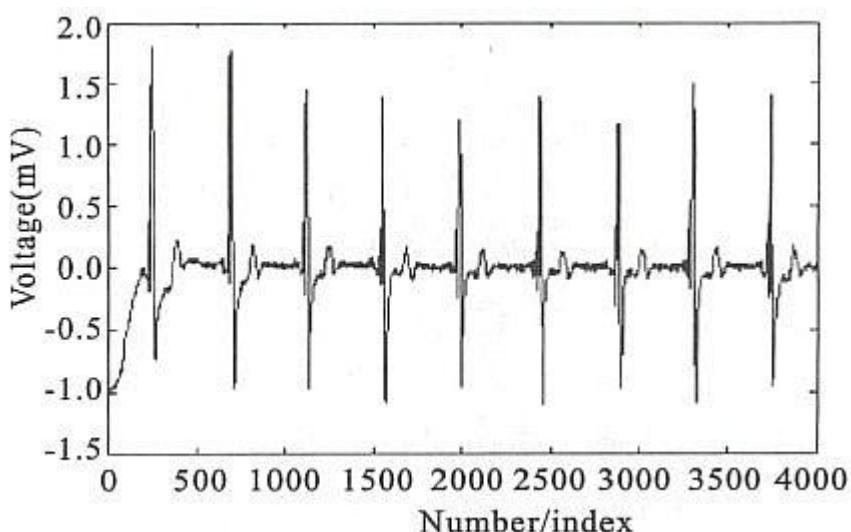


图 2.16 NLMS 自适应算法处理抑制 ECG 信号基线漂流

2.3 符号动力学概述

2.3.1 符号动力学基本理论

符号化动力学^[26]起源于动力系统的抽象拓扑理论的研究，20世纪初阿达马首次提出符号化动力学观点。三十年代符号动力学开始应用于变分学和微分几何中，莫尔斯和赫德隆在这一领域做出了巨大贡献。20世纪60年代起，符号化动力学进一步发展完善，并应用于一维映射研究领域。符号化动力学的一个典型应用就是可以很好的描述马蹄映射，这种映射过程具有迭代特征，是一个混沌系统，因此符号化动力系统是混沌系统的原型。符号化动力系统在离散时间动态系统控制问题研究领域也有重要应用。

符号化动力学主要应用于动力系统研究领域。动力系统是指系统状态可以表示为有限个符号的无穷序列，而任一状态点引出的运动轨道可由表示该状态的无穷序列通过简单的移位规则来确定。由于许多复杂动态系统均可以经过变换等价于动力系统，因此符号化动力学广泛的应用于各种动态系统的研究领域。

在生物电信号的非线性研究中，符号动力学与熵结合的分析方法占据了重要地位。符号化熵分析法对于数据量要求小，算法效率高，因此受到越来越多的关注。其主要原理是：首先将时间序列，比如生物电信号，在幅度域进行符号化，得到符号化时间序列，是由有限个符号构成的。再将符号化时间序列按一定规律进行编码，实质上是考察各种子串模式在某一

固定长度下时间序列中的出现概率^[26]，并对其求信息熵。尽管在符号化的过程中会丢失一些细节信息，但是能够大大的降低噪声的一些，而且同时保留原始动力系统的本质的大尺度特性，因此广泛的应用于动力系统分析领域。

2.3.2 符号化映射方法及符号信息熵

符号动力学分析方法是指把原始的时间序列转化成由若干个符号组成的时间序列然后进行分析这样一种“粗略”方法。Kurths 最早提出符号动力学信息熵。符号动力学的理论来源主要是信息学理论和动力学理论，最早是用来研究动力学系统复杂的行为，比如在对混沌系统的研究中采用符号动力学方法能够起到简化数据表示的目的，发展至今已运用于其他许多领域，诸如统计，通信和生物信息等。符号动力学方法对处理连续状态空间的系统也是很有用的，许多研究者在连续数据研究中利用符号转换也取得了卓有成效的结果。并作为符号动力学理论的正式提出。考虑一个时间序列 X , $X = \{x_0, x_1, x_2, \dots, x_i, \dots, x_N\}$ 。我们把该时间序列 X 转换成符号序列 $S = \{s_1, s_2, s_3, \dots, s_i, \dots, s_N\}$, $s_i \in A$ ($A = 0, 1, 2, 3$)。具体的转换依据方程 (2.1)。

$$s_i(x_i) = \begin{cases} 0: & u < x_i \leq (1+a)u \\ 1: & (1+a)u < x_i < \infty \\ 2: & (1-a)u < x_i \leq u \\ 3: & 0 < x_i \leq (1-a)u \end{cases} \quad (2.1)$$

这里 $i = 1, 2, 3, \dots$ 。 u 代表该时间序列的平均值， a 是一个特殊参数[27]，我们取 $a = 0.05$ 。我们发现 a 取 0.04 到 0.07 时都可得到相似的结果。如果 a 取小于 0.04 或大于 0.07 的值，则得不到较为合理的统计特性。这主要是因为在把原始时间序列符号化的过程中， a 的值取的过大或过小，都会导致细节信息的丢失，信号中的动态信息不能很好的捕捉到。在符号动力学分析方法中，关键的一步就是如何按照原始时间序列的量值确定相应的符号划分区域，然后把原始时间序列转化成符号序列。利用符号序列代替原始序列进行计算的一个最大的优势就是数值计算的速度大大提高了。统计学上针对离散系统的马尔可夫链[3]，以及分析复杂生物资料的信息学[4]。(有兴趣的读者还可参考作者 Kitchens 的文献[5]，其中在每一章节的最后都包含了简洁的历史发展的概要)。虽然在序列符号化转化过程中，原始时间序列会丢失一些细节信息，但是时间序列的动力学特性保留了下来[1,2]。Collet 和 Eckmann [9] 指出符号动力学完全能够捕捉和描述动力学系统的行为，符号动力学分析方法最大的特点就是运算简单、快速。

对于符号化序列 $\{s_i\}$ ，其所有子串的集合可以表示为 $\{W_j\}$ ($1 \leq j \leq N-m+1$)， $\{W_j\} = (s_j, s_{j+1}, \dots, s_{j+m-1})$ ，考察 $\{W_j\}$ 的分布特性，即统计出各子串在 $\{W_j\}$ 出现的概率 $p(l)$ ，从而得到对于的时间序列的符号化信息熵如方程 (2.2) 所示：

$$H_k = - \sum_{p(l)>0} p(l) \log p(l) \quad (2.2)$$

2.4 常用生物电数据库

近年来，为了对身体内部器官的健康状况进行准确的预报和诊断，而各种生物电信号能够正确的反应人体的健康状况，为临床医疗的预报和诊断提供依据。同时需要行之有效的方法来采集和分析人体生物电信号。自生物电图技术首次应用于临床诊断至今已有一个世纪的时间，不断发展和改进的生物电信号采集及相应分析技术为保护人类的生命与健康做出了卓越贡献。

鉴于生物电信号在科研、医疗以及社会方面的重要意义，标准生物电数据库建立和完善就变得极其重要。生物电数据库是指各种生物电图按照特定的分类进行存储，并按照一定的组织规律进行调用的数据库。必须按照严格的仪器设备标准、诊断标准和分类标准来建立生物电信号数据库。主要因为标准数据库主要应用于科研及医疗领域，均要求极高的精准性和科学性。只有标准的数据才能作为临床诊断的参考依据，提高医务工作者的诊断精度；而且可以检验和评价生物电设备的自主分析诊断能力。同时，标准生物电数据库在教学研究领域也有重要应用。生物电数据库主要由生物电数据集组成，只有经过临床证实的资料才可以入库，并且所有数据都经过医学专家的鉴别和标注。目前我国的各种生物电生物电信号数据库的建设尚处在初级阶段，信息化工作平台已经广泛应用于各医院的心脑疾病的诊断治疗，临床数据也在不断的采集和完善中，规范化、标准化生物电数据库的建设指日可待。

目前，在生物电研究领域中使用比较广泛的生物电数据库主要有四个：MIT-BIH 数据库、ST-T 心电数据库、CSE 心电数据库及 AHA 心率失常心电数据库。另外 PTB 心电数据库、Sudden Cardiac 数据库、PAF 预测数据库等心电数据库也在国际上得到广泛认可。各个 ECG 数据库中数据特点和侧重各不相同，分布针对不同的需求和使用要求，其中欧盟的 ST-T 心电数据库主要是辅助临床诊断；PTB 心电数据库则是针对教学科研以及心电设备校正；而 MIT-BIH 数据库中的数据比较全面，能够适应教学研究和辅助临床诊断等多种需求。上述数据库中除 AHA 心律失常心电数据库和欧盟 CSE 数据库需要购买外，其他的都可以从 PhysioNet 网站上自由下载。

MIT-BIH 数据库是应用最为广泛的心电数据库，是由美国麻省理工学院与 Beth Israel 医院联合建立的。它包括很多子数据库，每个子数据库包含特定类型的心电记录^[28]。从 1999 年起，MIT-BIH 数据库在 Internet 上公布，允许人们通过互联网免费下载和使用^[28]，这一行为得到了美国国家研究资源中心和国家健康研究院的支持，也大大推动了心电信号的科研发展。由于该数据库具有较高的权威性，常被作为实验数据的来源和各类识别算法的检测标准，国内外许多心电方面的研究都是基于该数据库的，其中应用最多的是 MIT-BIT QT 数据库和

MIT-BIT 心律不齐数据库。本文采用的实验数据也来自该数据库。

MIT-BIH 数据库主要包含以下子库：

- 1) MIT-BIH 心率失齐数据库，该数据库来源于心率失齐记录。其采样频率是 360Hz，分辨率 11bit，每段数据持续时间为 30 分钟，存储格式为 Format 212。
- 2) MIT-BIH ST 段变化数据库，其采样频率是 360Hz，分辨率为 12bit，每段数据持续时间不定，存储格式为 Format 212。
- 3) MIT-BIH 心房震颤数据库，其采样频率是 250Hz，分辨率为 12bit，每段数据持续时间 10 小时，存储格式为 Format 212。

AHA 心律失常心电数据库^[28]的采样频率是 250Hz，分辨率为 12bit。它由美国国家心肺及血液研究院资助的美国心脏协会开发，记录了心律失常心电数据，其主要应用是评价室性心律不齐探测器的检测效果。

欧盟的 ST-T 数据库是由欧洲心脏病学会开发的，主要应用于评价 ST 段和 T 波检验算法性能的数据库，其采样频率是 250Hz，分辨率为 12bit，每段数据持续时间 2 小时，存储格式为 Format 212。

PTB 心电诊断数据库是由德国国家计量署提供的数字化心电数据库，其目的在于算法标准的研究与教学。数据来源于柏林的本杰明富兰克林医学大学的心脏内科。其采样频率是 1000Hz，分辨率为 16bit，每段数据持续时间不定，存储格式为 Format 16。

PAF 预测挑战数据库来自于 2001 年的开放性竞赛，这个竞赛是针对自动预测阵发性心房纤颤/震颤的，竞赛的意义是刺激并促进美国在这个重大临床问题上的探索和培养友好竞争和广泛合作的环境。其采样频率是 128Hz，分辨率为 16bit，每段数据持续时间为 5 分钟或 30 分钟，存储格式为 Format 16。

心脏性猝死动态心电数据库是由 PhysioNet 建立的，主要应用于心脏性猝死这一领域的电生理研究。其采样频率是 250Hz，分辨率为 12bit，每段数据持续时间为 30 分钟，存储格式为 Format 212。

欧盟的 CSE 数据库包含了 1000 例短时间的心电记录，主要应用于评价心电图自动分析仪的性能。采用 12 或 15 导联。

论文主要采用的是第三个 ECG 数据库，即 MIT-BIH 数据库作为实验数据来源，主要是因为该数据库中的包含的 ECG 数据类型比较全面，数据量也比较大；该数据该库中的记录是多参数睡眠数据，包括 1 导 EEG 信号，1 导 ECG 信号，1 导 EOG 信号，1 导 EMG 信号等多导睡眠信号，记录长度为 6 个小时，数据采样频率 250Hz，每份记录的数据都附带着以 30s 为一个分期的注释信息。该数据库一共包含 48 条记录，每条记录的数据结构是 2 个通道的

ECG 数据，每条记录都是从 24 小时的 ECG 记录中经过两个以上心脏专家选出具有代表性的 30 多分钟的数据片段，并手工标注后形成；加上与之对应的记录头文件和标记文件。通道中 ECG 数据的采样率为 360Hz，量化精度为 11bit。在记录头文件中主要是保存了对应记录的具体描述诸如 ECG 数据格式、时间长度和病历等此类的信息。标注文件主要是包含了 ECG 记录所对应的心律和记录的 ECG 信号质量等信息。此外，MIT-BIH 数据库中的部分数据（主要是心律失常方面）和 PTB 心电数据库以及 European ST-T 心电数据库是相互共享数据的关系，都可以从 Physionet^{*}网站上得到。

2.5 小结

本章首先介绍了生物电活动的生物学基础，分别介绍了脑电、心电、肌电的生理机制。接着介绍各种生物电信号采集方法，并给出了生物电信号采集系统构成。常用的生物电信号采集系统一般由以下几个部分组成^[22]：提供输入阻抗、导联系统、前级放大、脱落检测、右腿驱动、多路选通、后级放大及滤波、信号量化、采集控制模块和工作站。然后指出了生物电中几种比较常见的干扰信号类型以及对应传统处理方法。干扰生物电信号的噪声主要有：工频干扰、基线漂移、肌电干扰、电磁设备干扰和电极接触干扰等。并在此基础之上，采用了自适应滤波对原始生物电信号进行降噪处理，来去掉主要的干扰—工频干扰和基线漂移。并简单介绍了符号化动力学理论。最后简介了常用生物电数据库极其基本参数，主要有 MIT-BIH 数据库、ST-T 心电数据库、CSE 心电数据库及 AHA 心率失常心电数据库。另外简单介绍了 PTB 心电数据库、Sudden Cardiac 数据库、PAF 预测数据库。

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

第三章 基于符号化相对熵的脑电信号时间不可逆性分析

3.1 熵增原理与时间不可逆性

熵用来表示任何一种能量在空间中分布的均匀程度。熵的概念由德国物理学家克劳修斯于1865年所提出。一个热力学系统中熵的增减定义如下：在一个可逆性程序里，被用在恒温的热的总数 δQ ，用公式表示为：

$$\Delta S = \frac{Q}{T} \quad (3.1)$$

熵代表着一个系统不受外部干扰时往内部最稳定状态发展的特性。熵函数是一种状态函数^[29]，又是一个判别习惯函数，能够定量说明自发过程的方向和限度。一个系统的熵在体系的能量完全均匀分布时达到最大值。克劳修斯认为熵是在学习可逆及不可逆热力学转换时的一个重要元素。一个系统能量分布的越均匀^[29]，熵就越大。

让两个温度不同的问题相互接触，那么两个物体间将产生能量传递^[29]，即冷物体将变热，热物体将冷却，直到两个物体到达相同的温度为止，如果任由一个系统自然发展，那么能量差总是倾向于消除的，这个过程即熵增过程。熵增原理是指系统的熵在可逆绝热过程中不变，在不可逆绝热过程中单调增大。即对于绝热过程($Q=0$)，故 $S \geq 0$ 。熵增过程是一个自发的从有序状态逐渐无序的过程，熵增过程是不可逆的^[29]，一个孤立系统的熵永远不会减少。熵增原理适用于孤立系统，这是由于孤立系统内部是绝热过程，其内部的一切变化与外界无关。孤立系统过程进行的方向和限度取决于熵的变化和最大值。在一个孤立系统由非平衡态趋于平衡态的过程中，其熵单调增大，当系统达到平衡态时，熵达到最大值。能量守恒定律适用于自然界，这是问题的关键，但是自然界没有完全孤立的系统，因此从整体是来看整个自然界是符合熵增原理的。

一切物质运动过程都具有的持续性和不可逆性构成了它们的共同属性，这种共同属性就叫做时间。所谓持续性，包括了过程的因果性和不间断性；所谓不可逆性，则指物质过程能重复但不可能返回过去的性质。从热力学角度而言，自然界所有的热力学过程都具有宏观上的不可逆性。这是由于在复杂热力学过程中，分子间的相互作用使得系统会在不同的热力学态之间演化。在演化过程中必然会有彼此做功的过程，在做功过程中产生热量的流失和耗散，而且这些能量的耗散是不可逆转的，即不能返回过去的性质。时间不可逆性观念的提出有重大的科学应用和价值：首先它从科学角度提出了新的时间观，热力学第二定律抓住了时间的

不可逆特性，确认了时间在物理世界演化中的尺度作用；其次提出了新的科学方法论，并拓展到可观测的宇宙起源；第三拓宽了科学研究领域和对象，延伸到探究人体的内在功能和机理。由于一切热力学变化均可以采用熵函数来对其进行分析，因为其方向和限度都可以归结为热和功转化及其限度的问题。

3.2 基于符号相对熵脑电信号时间不可逆性研究^[29]

3.2.1 相对熵理论

生命过程的不可逆性和熵产之间的关系是统计力学和物理学的核心^[30]。以物理学角度来理解生命。生命就是正向熵流和负向熵流的同时作用。自然界的一切自发进行的过程都是朝熵增的方向进行的。从统计力学角度来讲，一个生理机制产生的静态时间序列的具有时间逆不对称性，这种不对称性可以提供该生理机制的熵产信息^[31]。在熵增原理的前提下，物理学和统计力学紧密联系起来，实现了利用统计力学来量化研究生命现象。

熵的本质是变化的方向性和时间的方向性，熵指明了不可逆过程的方向性。当过程处于比较有序的状态时熵值较小，处于比较无序的状态时具有较大的熵值。可逆过程的熵值为零，现实中是不存在可逆过程的。过程的不可逆程度越大，熵产越大，即能量耗散越大^[32]。

如果把任意的可逆循环分割成许多小的卡诺循环，可得出：

$$\sum (\delta Q_i / T_i)_r = 0 \quad (3.2)$$

即任意的可逆循环过程的热温熵之和为零。其中， δQ_i 为任意无限小可逆循环中系统与环境的热交换量； T_i 为系统温度。

对于可逆过程，熵 S 与热温熵有以下关系：

$$\Delta S = \delta Q_r / T \quad (3.3)$$

对于不可逆过程，则有：

$$\Delta S > \delta Q_r / T \quad (3.4)$$

这就是克劳修斯不等式，表明在经历了一个微小不可逆变化后，该变化过程中的热温熵大于一个隔离系统的熵变。可逆过程的每一步微小变化，都无限接近于平衡状态。而不可逆过程是所有自发过程的共同特征，不可逆过程所能达到的限度是平衡状态即可逆状态，即公式(3.4)可以作为这一过程自发与否的判断依据。

本文在熵增原理的基础上，引入一种分析脑电信号时间不可逆特性的新方法——符号相

对熵^[1]。相对熵量化了熵增和不可逆性之间的关系，将物理学和统计力学联系起来^[33-35]。本文提出了一种新的使用过程的前向概率和后向概率计算符号相对熵的方法，利用符号相对熵来估计能量耗散。并运用该方法分别对老年及青年脑电信号^[36-40]进行处理，最后针对实验结果进行了分析讨论。

时间序列的不可逆性和熵产之间的关系是物理学和统计力学的核心，两个随机过程的概率分布 $p(x)$ 和 $q(x)$ 间的相对熵定义为：

$$D(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad (3.5)$$

相对熵是非负的，当且仅当 $p(x)$ 和 $q(x)$ 相同时才为零。但是 $D(q \parallel p) \neq D(p \parallel q)$ ，即相对熵不具有对称性^[1]。

熵产可用来衡量这两个随机过程间的可区别性。不同情况下的熵产和相对熵间的关系已经推导处理，也经过验证。

对于非平衡态，每个时间单位的熵产为：

$$\frac{\langle s \rangle}{k} = \lim_{t \rightarrow \infty} \frac{1}{t} D[p(\{x(\tau)\}_{\tau=0}^t) \parallel p(\{x(t-\tau)\}_{\tau=0}^t)] \quad (3.6)$$

其中 k 是波尔兹曼常数， $p(\{x(\tau)\}_{\tau=0}^t)$ 是一个给定轨迹 $\{x(\tau)\}_{\tau=0}^t$ 的概率分布。对于静态轨迹， $\tilde{x}(\tau)$ 唯一确定熵产 s 。

3.2.2 基于符号序列的相对熵方法

对于离散静态曲线 x_1, \dots, x_n ，本文提出了一种新的符号相对熵的计算方法^[1]，即定义长度为 n 的序列的符号相对熵为：

$$D_n(p_F \parallel p_B) \equiv \frac{1}{n} \sum_{l=1}^n p(\dot{x}_1, \dots, \dot{x}_n) \log \frac{p(\dot{x}_1, \dots, \dot{x}_l)}{p(\dot{x}_l, \dots, \dot{x}_n)} \quad (3.7)$$

这个公式揭示出物理机制和时间序列的统计学信息有密切联系^[1]。公式左侧是一个纯物理量，而右侧是仅与序列 x_1, \dots, x_n 有关的纯统计量。这种联系是 Landauer's 定律的泛化，Landauer's 定律将熵产和逻辑不可逆性联系起来。根据公式(3.7)，我们可以通过计算一个时间序列前向分布和后向分布间的相对熵，来得到产生该序列的生理机制的熵增信息，据此我们进而可以判断该生理过程是活跃的还是消极的。

主要流程如下：

(1) 原始序列符号化：

我们的目的就是寻找一种有效的方法，需要很少的数据量就可统计出有意义的结果，运算速度快，能够分析非平稳的且有噪声的数据，为分析实际的时间信号提供方便。目前符号映射方式有以下三种：一种是以整个时间序列的统计特性，例如均值或方差等，为基准作为符号状态空间划分的阈值，或直接界定符号空间的临界点，例如 Kurths 等提出的符号动力学信息熵中采用的符号映射方法；第二种则以时间序列的局域波动作为符号映射的依据，能够很好的体现时间序列的高频特性，例如李锦^[27]等在对 HRV 信号进行基本尺度熵计算时采用的序列符号化的方法；第三种则是以时间序列中的一阶差分序列作为符号映射的参考依据，例如卞春华^[42]等提出的符号序列熵。

本文根据公式 (3.8) 将原始序列符号化为只由 0,1,2,3 组成的序列：

$$s_i(x_i) = \begin{cases} 0 : \mu_1 < x_i \leq (1+a)\mu_1 & \text{or } (1+a)\mu_2 < x_i < \mu_2 \\ 1 : (1+a)\mu_1 < x_i < \infty & \text{or } -\infty < x_i < (1+a)\mu_2 \\ 2 : (1-a)\mu_1 \leq x_i < \mu_1 & \text{or } \mu_2 < x_i < (1-a)\mu_2 \\ 3 : (1-a)\mu_2 < x_i < (1-a)\mu_1 \end{cases} \quad (3.8)$$

(2) 对符号化序列进行编码——统计字串出现的概率

设符号化得到的序列为 $S = \{s_0, s_1, s_2, \dots, s_i, \dots, s_N\}$ ，其中 $s_i = 0, 1, 2, 3$ ；那么根据公式 (3.9) 将符号化序列编码。

$$bx_i = s_i \times 4^2 + s_{i+1} \times 4^1 + s_{i+2} \times 4^0 + 1 \quad (3.9)$$

其中 $i = 1, 2, \dots, N-2$ 。

例如符号序列为 30210313…，则依次组合得到 302, 021, 210, 103, 031, 313…。由 0、1、2 和 3，这四个符号组成的三位字状态空间中共有 $4^3 = 64$ 种不同状态；如三位字 302，编码值为 $3 \times 4^2 + 0 \times 4^1 + 2 \times 4^0 + 1 = 51$ 。因此符号序列为 30210313 编码得到的编码序列为 51, 10, 37, 20, 14, 56。

(3) 根据公式 (3.7) 逐点求前向概率和后向概率间的相对熵，并取平均值，得到单位时间上的平均能量耗散。

我们对序列逐点向前向后求其概率分布^[1]，以序列 x_1, \dots, x_n ，为例，对于其中的每一点 x_i ($1 \leq i \leq n$)，以 x_i 为界将原序列分为两个序列，其中 p_{F_i} 为序列每一点向前的概率分布， p_{B_i} 为序列每点向后的概率分布，计算前向概率 p_{F_i} 和后向概率 p_{B_i} 间的相对熵 $D_i(p_{F_i} || p_{B_i})$ 。

我们对序列逐点求 $D_i(p_{F_i} || p_{B_i})$, 最后对其进行平均, 得到该时间序列的符号化相对熵, 它代表了序列的不可逆程度。由生理机制驱动产生一个随机过程, 那么该生理活动的平均能量耗散就定义为描述这个随机过程的时间前向概率分布和后向概率分布间的符号相对熵。符号相对熵反映了随机过程前向分布和后向分布的差异性。相对熵越大, 差异性越大。序列的不可逆性也就越大。

算法流程如图 3.1 所示:



图 3.1 算法基本流程框图

熵增原理将熵产和逻辑不可逆性联系起来^[1], 公式(3.8)是对熵增原理的扩展。根据本章提出的算法, 我们首先对脑电信号进行相应的符号化及编码处理后, 再利用公式(3.7)计算脑电序列的相对熵, 然后根据公式(3.8)计算出脑电信号的熵产生率, 即平均能量耗散。据此我们可以判断该脑电好的是积极的还是消极的, 这对于脑类疾病的诊断和资料有着重大意义。计算过程包括符号化、编码、逐点求序列前向概率及后向概率、计算符号相对熵和熵产生率, 最后使用 spss 统计分析软件对计算结果进行了假设检验, 验证了该算法的有效性。

3.3 数据分析

3.3.1 EEG 实验数据

我们从 Neuroelectric and Myoelectric Databases 数据库中随机提取十位健康年轻人及十位健康老年人的脑电信号。其中年轻人数据分别取自数据库中的“yng”文件(包括 fMT1419, fMT1420, fMT1415, fMT1416, fMT1427, fMT1421, fMT1422, fMT1423, fMT1424, fMT1428), 记为样本 YngEEG; 老年人数据取自数据库中的“eh”文件(包括 fMT1502, fMT1503, fMT1504, fMT1506, fMT1523, fMT1510, fMT1509, fMT1512, fMT1519, fMT1520), 记为样本 OldEEG。

3.3.2 实验方法与结果分析

对受训者样本 YngEEG 和 OldEEG 脑电信号进行相对熵分析, 首先对这两组非线性数据序列进行符号化, 编码, 相空间重构后, 选择数据长度为 2000 分别计算其相对熵, 并进行显著性检测, 然后进行多样本验证。

A.研究平均能量耗散与数据长度 L 的关系

对受训者样本 YngEEG 和 OldEEG 中的每组脑电信号序列分别取数据长度 $L=400, 800, 1200, 1600, 2000, 2400$, 计算每个个体的能量耗散, 并分别对青年样本和老年样本进行平均, 得到平均能量耗散与数据长度 L 的关系如表 3. 所示(编码长度 N=8) :

表 3.1 (a) 青年人能量耗散与数据长度 L 的关系

数据长度	L=400	L=800	L=1200	L=1600	L=2000	L=2400
个体 1	16.649	23.113	22.158	31.228	34.872	31.053
个体 2	15.585	16.902	10.009	32.649	33.35	27.173
个体 3	2.766	9.8955	11.664	30.838	23.538	21.355
个体 4	13.16	7.3708	20.187	30.37	35.573	38.685
个体 5	3.5205	18.756	18.6	23.626	28.211	30.696
个体 6	7.158	12.79	21.988	21.207	30.915	29.512
个体 7	8.8535	23.415	28.428	33.337	27.761	31.462
个体 8	8.0133	17.47	31.691	33.207	27.014	34.75
个体 9	4.256	19.142	47.487	22.164	24.568	43.482
个体 10	7.2406	4.1293	21.873	14.002	42.449	29.841
平均值	8.7202	15.298	23.409	27.263	30.825	31.801

表 3.1 (b) 老年人能量耗散与数据长度 L 的关系

数据长度	L=400	L=800	L=1200	L=1600	L=2000	L=2400
个体 1	7.813	10.544	15.945	43.333	48.818	44.826
个体 2	10.099	20.813	36.057	32.297	32.15	38.737
个体 3	17.94	37.65	32.725	41.06	50.992	28.093
个体 4	18.176	24.557	24.963	36.345	30.376	26.326
个体 5	7.8844	11.335	8.7077	15.934	12.133	11.533
个体 6	4.7678	21.693	26.034	23.792	33.016	37.313
个体 7	19.425	32.798	35.612	37.136	36.762	36.851
个体 8	15.087	31.691	43.715	59.542	85.437	111.81
个体 9	7.9245	11.521	19.976	45.593	16.862	24.402
个体 10	17.932	15.461	28.329	24.613	22.166	38.386
平均值	12.7049	21.8063	27.2064	35.9645	36.8712	39.828

青老年平均能量耗散的直接对比关系如图 3.2 所示：

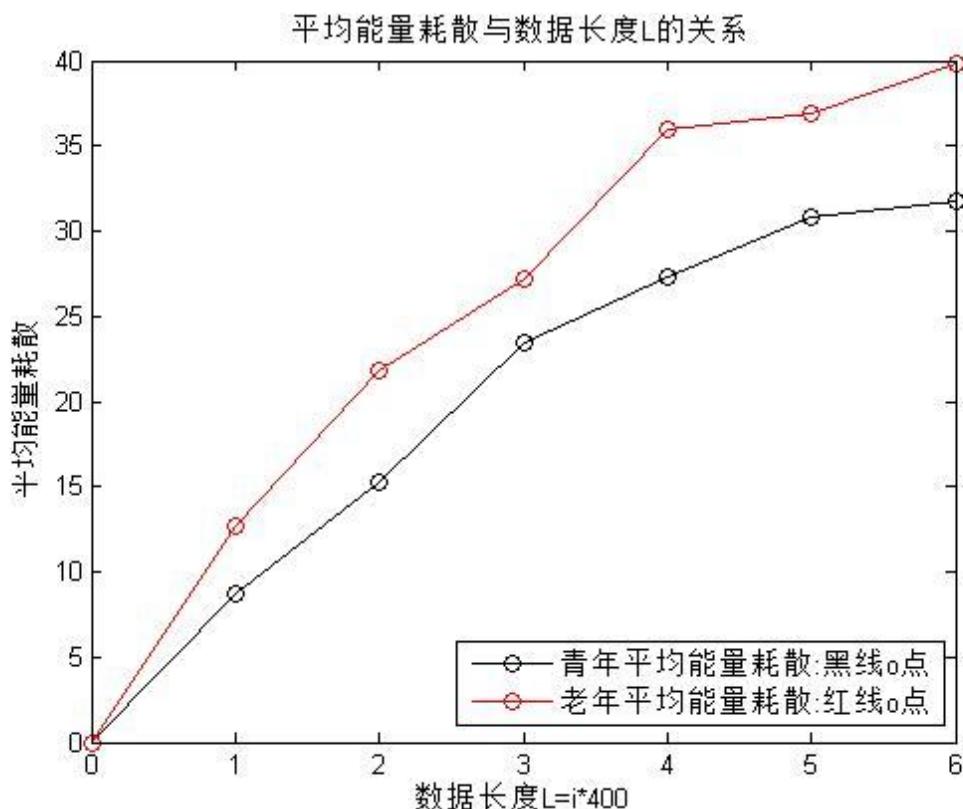


图 3.2 平均能量耗散与数据长度 L 的关系

分析图 3.2 青老年能量耗散曲线可知^[1]:

- 1) 在数据长度相同的情况下，老年人的平均能量耗散要高于青年人。
- 2) 当数据长度由 0 增至 1600 过程中，青老年的平均能量耗散值也呈递增趋势；数据长度大于 1600 时，平均能量耗散值趋于平稳，基本不变。

从数据的精准性考虑，数据长度越大，其统计概率分布越接近实际分布，相应的准确性越高。但从计算量以及计算速度来讲，数据长度越小，那么算法速度也越快，综合我们的实验结果，取数据长度 $L=1600$ 时，即能兼顾处理速度，又可以保证实验精度。

B.研究平均能量耗散与编码长度 N 的关系

对受训者样本 YngEEG 和 OldEEG 中的每组脑电信号序列分别取编码长度 $N=1, 2, \dots, 20$ ，计算每个个体的能量耗散，并分别对青年样本和老年样本进行平均，得到平均能量耗散与编码长度 N 的关系如表 3.2 所示（数据长度 $L=500$ ，限于表格大小仅保留两位小数）：

表 3.2 (a) 青年群体能量耗散与编码长度 N 的关系

编码 长度	个体 1	个体 2	个体 3	个体 4	个体 5	个体 6	个体 7	个体 8	个体 9	个体 10
N=1	5.41	6.32	9.19	7.01	9.24	6.25	5.65	7.28	6.28	7.64
N=2	7.16	6.52	15.39	9.16	20.46	13.58	7.69	13.01	10.08	12.90
N=3	11.81	7.30	17.41	12.15	24.07	15.82	8.27	19.04	11.23	20.27
N=4	11.42	7.61	18.38	13.61	25.76	16.84	8.47	23.60	11.73	25.07
N=5	11.15	8.93	18.92	15.84	26.69	17.25	8.52	26.96	11.94	27.49
N=6	10.93	9.78	19.25	15.95	27.24	17.38	8.50	29.44	12.03	28.02
N=7	10.74	10.29	19.45	16.02	27.50	17.34	8.42	32.00	12.03	28.33
N=8	10.57	10.81	19.58	16.07	27.07	17.20	8.33	32.96	11.98	28.53
N=9	10.41	10.46	19.65	15.93	26.70	16.99	8.22	33.02	11.89	29.29
N=10	10.26	10.16	19.69	15.81	26.36	16.74	7.95	32.82	11.78	28.49
N=11	10.12	9.88	19.70	15.71	26.06	16.38	7.67	32.61	11.66	27.83
N=12	9.97	9.63	19.69	15.61	25.64	16.04	7.41	32.04	11.51	27.25
N=13	9.83	9.38	19.66	15.10	24.95	15.69	7.17	31.50	11.36	26.73
N=14	9.69	9.15	19.63	14.66	24.34	15.32	6.93	30.99	11.20	26.07
N=15	9.56	8.93	19.58	14.27	23.79	14.95	6.71	30.87	11.03	25.48
N=16	9.42	8.71	19.53	13.93	23.30	14.57	6.49	30.41	10.86	24.94
N=17	9.28	8.51	19.48	13.63	22.73	14.18	6.28	29.97	10.68	24.44
N=18	9.15	8.30	19.42	13.35	22.02	13.79	6.07	29.54	10.49	23.99
N=19	9.01	8.10	19.38	13.10	21.37	13.41	5.86	29.12	10.30	23.56
N=20	8.88	7.90	18.99	12.87	20.78	13.04	5.66	28.48	10.11	23.16

表 3.2 (b) 老年群体能量耗散与编码长度 N 的关系

编码 长度	个体 1	个体 2	个体 3	个体 4	个体 5	个体 6	个体 7	个体 8	个体 9	个体 10
N=1	8.29	6.59	5.84	6.08	6.63	5.37	10.28	8.71	5.44	6.67
N=2	15.29	7.85	8.85	7.43	5.41	5.08	10.69	12.41	9.29	5.48
N=3	17.59	8.43	9.75	8.72	5.02	4.99	10.85	13.61	11.24	5.10
N=4	18.71	8.81	10.12	10.02	4.84	4.95	10.93	14.18	12.72	4.93
N=5	19.34	8.94	10.26	11.98	4.74	4.91	10.99	14.49	12.59	4.83
N=6	19.74	8.65	10.30	14.31	4.67	4.88	11.02	14.66	12.45	4.76
N=7	20.00	7.97	10.27	15.09	4.63	4.85	11.05	14.76	12.30	4.71
N=8	20.17	7.44	10.20	14.96	4.59	4.82	11.07	14.82	12.14	4.68
N=9	20.28	6.99	10.11	14.50	4.56	4.79	11.09	14.84	11.98	4.65
N=10	20.36	6.68	9.99	14.11	4.54	4.76	11.11	14.84	11.82	4.63
N=11	20.40	6.35	9.87	13.77	4.52	4.73	11.14	14.82	11.65	4.61
N=12	20.42	6.01	9.73	13.46	4.51	4.70	11.19	14.79	11.49	4.60
N=13	20.43	5.70	9.59	13.18	4.49	4.67	10.62	14.75	11.32	4.58
N=14	20.36	5.42	9.35	12.92	4.48	4.64	10.14	14.71	11.14	4.57
N=15	20.14	5.15	9.09	12.69	4.47	4.60	9.72	14.66	10.97	4.56
N=16	19.93	4.91	8.85	12.34	4.46	4.57	9.35	14.62	10.80	4.55
N=17	19.74	4.68	8.61	11.96	4.44	4.54	9.02	14.59	10.62	4.54
N=18	19.57	4.46	8.39	11.61	4.43	4.50	8.73	14.17	10.44	4.53
N=19	19.40	4.25	8.17	11.17	4.42	4.47	8.46	13.79	10.26	4.52
N=20	19.24	4.05	7.96	10.77	4.41	4.43	8.23	13.43	10.08	4.51

从表 3.2 中可以看出^[1], 无论是青年个体还是老年个体, 在编码长度 N 从 1 递增时, 其能量耗散先是呈递增趋势。当 N=8 时基本趋于平缓。当 N 大于 14 时开始缓慢减小, 基本不变。

表 3.3 青老年脑电信号平均能量耗散对比

编码长度	青年人 平均值	老年人 平均值	青年人 最小值	老年人 最小值	青年人 最大值	老年人 最大值
N=1	7.028867	6.989325	5.411252	5.368509	9.239364	10.27664
N=2	11.59388	8.777441	6.516041	5.08127	20.46109	15.29198
N=3	14.73813	9.532043	7.304709	4.992658	24.06778	17.59111
N=4	16.25047	10.01957	7.605124	4.843345	25.76346	18.70587
N=5	17.37066	10.30629	8.524234	4.740078	27.49416	19.34319
N=6	17.85051	10.54405	8.496578	4.673445	29.43685	19.73999
N=7	18.21364	10.56334	8.424551	4.626737	32.00487	19.99837
N=8	18.3085	10.48895	8.326616	4.591879	32.9568	20.16973
N=9	18.25621	10.37991	8.218258	4.564534	33.02088	20.28286
N=10	18.00642	10.28478	7.952599	4.542189	32.81843	20.35526
N=11	17.76169	10.18657	7.674282	4.523301	32.61441	20.39841
N=12	17.47976	10.08941	7.414517	4.506875	32.03893	20.42065
N=13	17.13848	9.934257	7.16884	4.492248	31.50114	20.43021
N=14	16.79966	9.772976	6.934087	4.47896	30.99283	20.36435
N=15	16.51699	9.605465	6.707963	4.466687	30.86834	20.13959
N=16	16.21625	9.436373	6.488767	4.455193	30.41217	19.93381
N=17	15.91688	9.273943	6.275227	4.444304	29.97078	19.74361
N=18	15.61175	9.081888	6.066378	4.43389	29.54165	19.56635
N=19	15.32203	8.891402	5.861495	4.247937	29.12287	19.39996
N=20	14.9877	8.712705	5.660043	4.05181	28.47662	19.24281

青老年脑电信号平均能量耗散更直观的对比如图 3.3 所示：

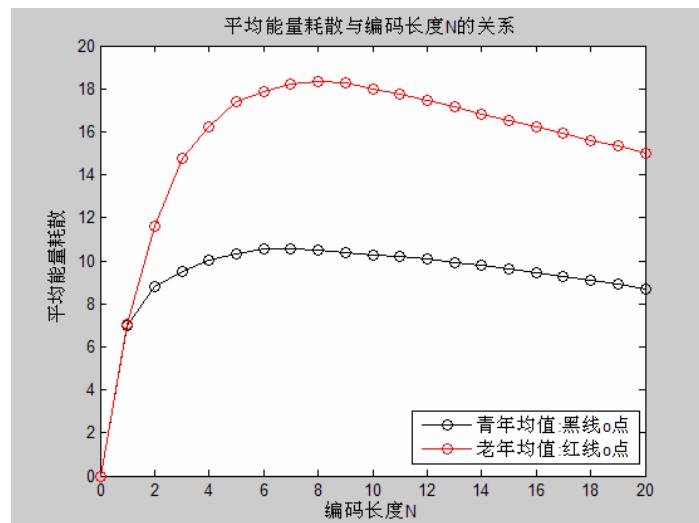


图 3.3 平均能量耗散与编码长度 N 的关系

分析图 3.3 平均能量耗散与编码长度 N 的关系可知^[1]:

1) 在编码长度 N 从 1 到 8 的递增过程中, 青老年的平均能量耗散都成迅速增大趋势。从 N 取 7 之后平均能量耗散缓慢减小^[1]。

2) 老年人的平均能量耗散明显高于青年人的, 且在 N=8 时青老年差异最显著。

考虑到实验效果的明显性, n 应该取大些; 考虑到算法的复杂度会影响实验处理速度以及在临床应用上的实时性,n 应该取小一些.综合考虑上述因素, n 取 8。

C.研究在数据长度 L=500,编码长度 N=8 条件下青老年能量耗散的差异

对受训者样本 YngEEG 和 01dEEG 中的每组脑电信号序列分别取编码长度 N=8 及数据长度 L=500, 计算每个个体的能量耗散, 并分别对青年样本和老年样本进行平均, 得到平均能量耗散与年龄的关系如图 3.4 所示:

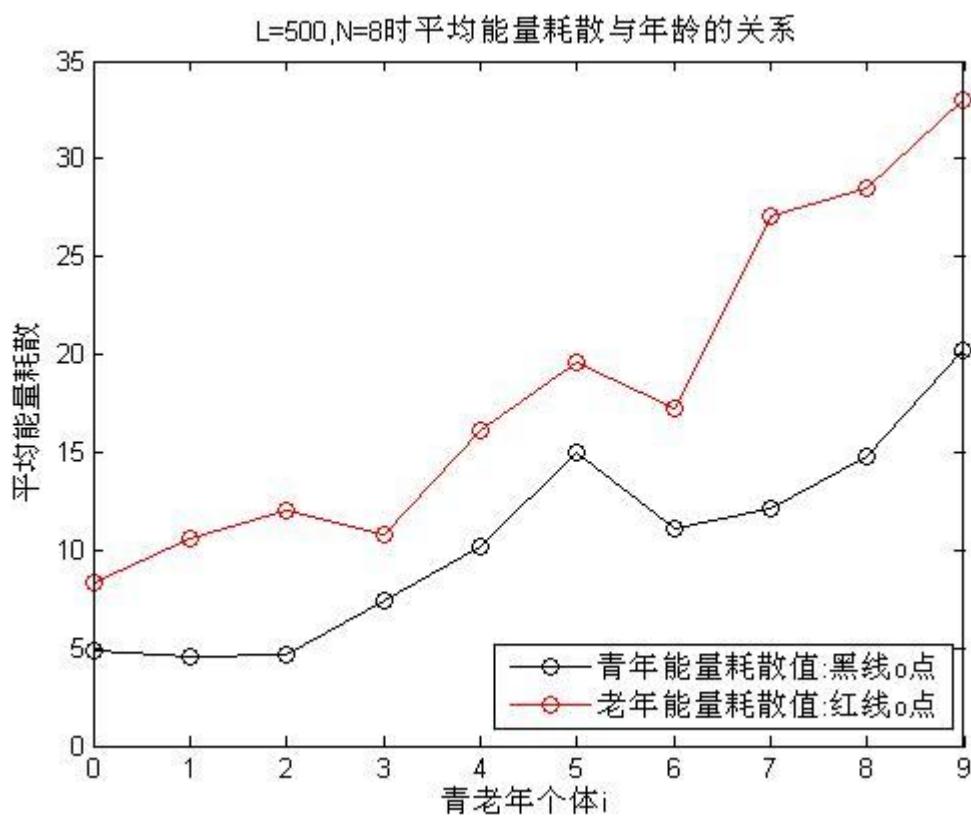


图 3.4 平均能量耗散与年龄的关系

分析图 3.4 平均能量耗散与年龄的关系可知: 老年人的能量耗散明显高于青年人。

D.统计分析与假设检验

为进一步验证本文计算结果的准确性及算法的有效性, 使用 spss 统计分析软件对计算结果进行了假设检验。主要方法如下: 对青老年平均能量耗散值的差异显著性进行假设检验, 使用 spss 对两组样本数据 Y、T 进行独立样本 T 检验, 结果如下表所示:

表 3.4 青老年平均能量耗散值差异显著性的独立样本 T 检验-组统计量

组统计量					
样本	分组	N	均值	标准差	均值的标准误
老年 青年	0	10	18.3085	8.559965	2.706899
	1	10	10.48895	5.222991	1.651655

表 3.5 青老年平均能量耗散值差异显著性的独立样本 T 检验-独立样本检验

独立样本检验					
样本		差方程的 Levene 检验		均值方程的 t 检验	
		F	Sig.	t	df
老年 青年	假设方差相等	2.942	0.103	2.466	18
	假设方差不相等	2.942	0.103	2.466	14.886

青老年平均 能量耗散值差异		均值方程的 t 检验			
		均值差值	标准误差值	差分的 95% 置信区间	
				下限	上限
老年 青年	假设方差相等	7.819557	3.171004	1.157526	14.481589
	假设方差不相等	7.819557	3.171004	1.157526	14.481589

根据表 3.5 分析青老年平均能量耗散差异的显著性，首先假设二者没有差异,分析表 3.5 可以得到以下结论：

①由表 3.5 可知，在假设方差相等的情况下，自由度(df)为 18，查找 t 值表可知理论 t 值 $t(df)_{0.05}=t(18)=2.101$ ，而样本 t 值为 2.466，大于 $t(18)$ ，且 $Sig(\text{双侧})=0.024<0.05$ ，假设不成立，所以青老年脑电信号平均能量耗散水平差异显著。

②由表 3.5 可知，在假设方差不相等的情况下，自由度(df)为 14，查找 t 值表可知理论 t 值 $t(df)_{0.05}=t(14)=2.145$ ，而样本 t 值为 2.466，大于 $t(14)$ ，且 $Sig(\text{双侧})=0.026<0.05$ ，假设不成立。所以青老年脑电信号平均能量耗散水平差异显著^[1]，证明该算法可以有效的对青老年脑电信号进行区分。

由前言中讨论可知，在一个孤立系统内部自然发生的过程总使系统不可逆的趋于熵取极大值的平衡态。将人的生命视作一个过程时,随着年龄的增长，生命的不可逆程度越大，熵产越大，即能量耗散越大。熵值较小对应于比较有序的状态，即年轻状态，也就是脑电活动的积极状态；熵值较大对应于比较无序的状态，也就是老年人脑电活动的消极状态。当生命体达到无序平衡高度混乱状态，熵增到最大值，生命衰老终结，那就是死亡。这与我们的实验

结果也是一致的，证明使用相对熵来计算过程不可逆程度的算法是可行的，我们可以以平均能量耗散作为特征值来衡量脑电信号是否处于积极状态。

3.4 小结

本文首先介绍了时间不可逆性研究现状。接着介绍基于符号相对熵的脑电信号时间不可逆性研究方法，并给出了相对熵理论和基于符号序列的相对熵方法。根据本章提出的算法，我们首先对脑电信号进行相应的符号化及编码处理后，计算过程包括符号化、编码、逐点求序列前向概率及后向概率、计算符号相对熵和熵产生率^[1]，最后使用 spss 统计分析软件对计算结果进行了假设检验，验证了该算法的有效性。主要的工作内容如下：

- (1) 基于过程不可逆特性提出了一种新的符号相对熵的计算方法^[1]，即使用过程的前向概率和后向概率来计算相对熵，从而获得过程的不可逆程度信息。
- (2) 根据熵增原理，通过相对熵得到过程的能量耗散，将物理学和统计学紧密联系起来，实现了利用统计学来量化研究生命现象。
- (3) 基于该算法对青老年人的脑电信号进行了分析，结果表明，老年人的能量耗散明显高于青年人，即老年人脑电活动的无序性也就是不可逆性显著高于青年人。该算法的计算结果与实际情况相符，证明该算法是正确的，表明能量耗散可以作为衡量一个过程是否处于积极状态的参数。

第四章 基于符号化相对熵的癫痫异常脑电信号分析

4.1 癫痫异常脑电信号病理机制及研究现状

4.1.1 癫痫异常脑类疾病病理机制及研究意义

癫痫（epilepsy）即俗称的“羊角风”或“羊癫风”^[41]，癫痫是一种意识、感觉、植物神经功能及精神异常的疾病，是由脑部神经元群阵发性异常放电引起的发作性运动，致病因素有很多。据中国最新流行病学资料显示，国内癫痫的总体患病率为 7.0%，年发病率为 28.8/10 万，1 年内有发作的活动性癫痫患病率为 4.6%。据此估计中国有 900 万左右的癫痫患者，其中 500~600 万是活动性癫痫患者，同时每年新增加癫痫患者约 40 万，在中国癫痫已经成为神经科仅次于头痛的第二大常见病。

癫痫病作为一种慢性疾病^[41]，虽然短期内对患者没有多大的影响，但是长期频繁的发作可导致患者的身心、智力产生严重影响。患者可能经受生命和精神上的危害。癫痫患者经常会在任何时间、地点、环境下且不能自我控制地突然发作，容易出现摔伤、烫伤、溺水、交通事故等；癫痫患者经常被社会所歧视，在就业、婚姻、家庭生活等方面均遇到困难，患者精神压抑，身心健康受到很大影响。而且存在认知障碍，主要表现为患者记忆障碍、智力下降、性格改变等，最后逐渐丧失工作能力甚至生活能力。严重者常并发急性肾功能衰竭，急性早幼粒细胞白血病，精神抑郁等症状。由于异常放电的起始部位和传递方式的不同，癫痫发作的临床表现复杂多样^[42-44]，可表现为发作性运动、感觉、自主神经、意识及精神障碍。引起癫痫的病因多种多样。经过正规的抗癫痫药物治疗，癫痫患者可以和正常人一样地工作和生活。临床统计数据表明约 70% 的癫痫发作是可以得到控制的，经 2~5 年的治疗 50%~60% 的患者可以痊愈。

一、癫痫的病因^[42]

癫痫病因复杂多样，包括遗传因素、脑部疾病、全身或系统性疾病等。一般认为与下列四种因素有关：

1) 遗传因素。遗传因素是导致癫痫尤其是特发性癫痫的重要原因。在一些有癫痫病史家族中容易出现癫痫。分子遗传学研究发现，一部分遗传性癫痫的分子机制为离子通道或相关分子的结构或功能改变。

2) 脑部疾病。包括先天性脑发育异常、颅脑肿瘤、颅内感染、颅脑外伤、脑血管病、变

性疾病、脑损害与脑损伤、胚胎发育不良及产伤等都可能会引起癫痫。

- 3) 环境因素；精神刺激、发热等也容易引起癫痫，另外农村发病率高于城市。
- 4) 全身或系统性疾病。比如缺氧、代谢性疾病、内分泌疾病、心血管疾病、中毒性疾病以及其他血液系统疾病等因素。

研究表明，癫痫病因与年龄的关系较为密切，不同的年龄组往往有不同的病因范围，如表 4.1 所示：

表 4.1 不同的年龄组常见病因

年龄组	癫痫诱因
新生儿及婴儿期	先天以及围产期因素（缺氧、窒息、头颅产伤）、遗传代谢性疾病、皮质发育异常所致的畸形等
儿童以及青春期	特发性（与遗传因素有关）、先天以及围产期因素（缺氧、窒息、头颅产伤）、中枢神经系统感染、脑发育异常等
成人期	头颅外伤、脑肿瘤、中枢神经系统感染性因素等
老年期	脑血管意外、脑肿瘤、代谢性疾病、变性病等

二、症状及分类^[43]

根据引起癫痫的病因不同，可以分为特发性癫痫综合征、症状性癫痫综合征以及可能的症状性癫痫综合征。2001 年国际抗癫痫联盟提出的新方案还对一些关键术语进行了定义或规范，包括反射性癫痫综合征、良性癫痫综合征、癫痫性脑病。根据病因不同可以癫痫分为以下类别：

- 1) 特发性癫痫综合征：除了癫痫，没有大脑结构性损伤和其他神经系统症状与体征的综合征。多在青春期前起病，预后良好。
- 2) 症状性癫痫综合征：是一种中枢神经系统病变或者异常，由于各种原因造成的，包括影响脑功能或者脑结构异常的各种因素。随着医学的进步和检查手段的不断发展和丰富，能够寻找到病因的癫痫病例越来越多。
- 3) 隐源性癫痫或可能的症状性癫痫综合征：指目前病因未明，但认为是症状性癫痫综合征的病症。
- 4) 反射性癫痫综合征：指几乎所有的发作均由特定的感觉或者复杂认知活动诱发的癫痫，如阅读性癫痫、惊吓性癫痫、视觉反射性癫痫、热浴性癫痫、纸牌性癫痫等。去除诱发因素，发作也消失。
- 5) 良性癫痫综合征：指不留后遗症的癫痫综合征，这种病症不需要治疗也能完全缓解，

或者易于治疗。

6) 癫痫性脑病：指癫痫性异常本身造成的进行性脑功能障碍。其原因主要或者全部是由于癫痫发作或者发作间歇期频繁的癫痫放电引起。大多为新生儿、婴幼儿以及儿童期发病。脑电图明显异常，药物治疗效果差。包括 West 综合症、LGS、LKS 以及大田原综合症、Dravet 综合征等。

按照发作方式不同癫痫又有不同分类。在 1981 年，癫痫发作分类方案由国际抗癫痫联盟提出，目前普遍应用的是该方案。癫痫发作根据该方案可分为全面/泛化性发作、部分性/局灶性发作、不能分类的发作等类型。2010 年国际抗癫痫联盟提出了最新的癫痫发作分类方案，新方案对癫痫发作进行了重新分类和补充。根据新方案临床发作类型有以下几种：

1) 全身强直-阵挛发作(大发作)：以突发意识丧失和全身强直和抽搐为特征，典型的发作过程可分为强直期、阵挛期和发作后期。一次发作持续时间一般小于 5 分钟，常伴有舌咬伤、尿失禁等，并容易造成窒息等伤害。强直-阵挛性发作可见于任何类型的癫痫和癫痫综合征中。

2) 失神发作(小发作)：以突发意识丧失和全身强直和抽搐为特征，典型的发作过程可分为强直期、阵挛期和发作后期。一次发作持续时间一般小于 5 分钟，常伴有舌咬伤、尿失禁等，并容易造成窒息等伤害。强直-阵挛性发作可见于任何类型的癫痫和癫痫综合征中。

3) 单纯部分性发作：发作时意识清楚，持续时间数秒至 20 余秒，很少超过 1 分钟。根据放电起源和累及的部位不同，感觉性、运动性、自主神经性和精神性都是单纯部分性发作的表现，后两种症状经常伴随出现，常发展为复杂部分性发作。

4) 复杂部分性发作(精神运动性发作)：表现为突然动作停止，两眼发直，叫之不应，不跌倒，面色无改变。有些患者可出现自动症，为一些不自主、无意识的动作，如舔唇、咂嘴、咀嚼、吞咽、摸索、擦脸、拍手、无目的走动、自言自语等，发作过后不能回忆。其大多起源于颞叶内侧或者边缘系统，但也可起源于额叶。

5) 继发性癫痫：复杂或简单部分性发作均可引起继发全面性发作。与全面性发作相比，部分性发作继发全面性发作在病因、治疗方法及预后等方面明显不同，仍属于部分性发作的范畴，故在临幊上其要重视这两者的鉴别。

6) 植物神经性发作：有晕厥型、腹痛型、头痛型、心血管性或肢痛型发作。

7) 强直发作：表现为发作性全身或者双侧肌肉的强烈持续的收缩，肌肉僵直，使肢体和躯体固定在一定的紧张姿势，如轴性的躯体伸展背屈或者前屈。常持续数秒至数十秒，但是一般不超过 1 分钟。强直发作多发于儿童患者，一般为病情严重的标志，有弥漫性器质性脑损害的癫痫患者常为该种类型，如 Lennox-Gastaut 综合征。

8) 肌阵挛发作：是肌肉突发快速短促的收缩，表现为类似于躯体或者肢体电击样抖动，

有时可连续数次，多出现于觉醒后。可为全身动作，也可以为局部的动作。肌阵挛临床常见，但并不是所有的肌阵挛都是癫痫发作。既存在生理性肌阵挛，又存在病理性肌阵挛。同时伴EEG多棘慢波综合的肌阵挛属于癫痫发作，但有时脑电图的棘慢波可能记录不到。一些预后较好的特发性癫痫患者（如少年肌阵挛性癫痫、婴儿良性肌阵挛性癫痫）可能属于肌阵挛发作；像早期肌阵挛性脑病、Lennox-Gastaut综合征、婴儿重症肌阵挛性癫痫等一些预后较差的癫痫综合征也可能属于该种类型，该种情况通常伴有弥漫性脑损害。

9) 痉挛：指婴儿痉挛，表现为突然、短暂的躯干肌和双侧肢体的强直性屈性或者伸性收缩，多表现为发作性点头，偶有发作性后仰。其肌肉收缩的整个过程大约1~3秒，常成簇发作。常见于West综合征，其他婴儿综合征有时也可见到。

三、病理生理和发病机制^[44]

癫痫的发病机制非常复杂。神经胶质细胞及离子通道神经递质发生改变时，会导致中枢神经系统兴奋与抑制间产生不平衡，从而引起癫痫发作。主要分为以下几类：

1) 离子通道功能异常

体内可兴奋性组织兴奋性调节的基础是离子通道，电压门控氯离子通道和钠离子通道是由两种基因独立编码控制。目前研究认为某些遗传性疾病的发生是由于离子通道功能受损，而离子通道功能受编码基因突变的影响。众所周知，人体内存在携带各种电位的离子，大量离子在穿过细胞膜的过程中会使得细胞电位发生改变。各种离子流形成离子通道，不同离子通道对细胞膜的作用不同：钠离子通道的主要作用是使细胞膜产生动作，这是通过兴奋细胞膜实现的。氯离子通道主要作用是稳定静息电位，这是通过抑制细胞膜的兴奋实现的。而目前认为很多人类特发性癫痫是由由缺陷的基因编码造成的有缺陷的离子通道病。研究表明，癫痫的发病与钙离子、钾离子、钠离子通道具有较强的相关性。

2) 神经递质异常

神经递质大多都会对患者的行为认知产生影响，正常情况下兴奋性与抑制性神经递质保持平衡状态，神经元膜稳定。癫痫性发病机制与神经递质关系极为密切，当兴抑制性神经递质不足、兴奋性神经递质过多或二者不平衡，都会使膜不稳定并产生癫痫性放电。

3) 神经胶质细胞异常

神经胶质细胞对维持神经元的生存环境起着重要的作用，这是由于神经元微环境的电解质平衡是维持神经元正常兴奋性的基础。神经胶质瘤细胞释放过多的谷氨酸盐可在肿瘤周围神经元产生癫痫活性。癫痫发作可能由星形胶质细胞对 γ 氨基丁酸或谷氨酸的摄取能力的变化引起。

四、临床表现及诊断^[45]

癫痫可见于各个年龄段。儿童癫痫发病率较成人高，随着年龄的增长，癫痫发病率有所降低。进入老年期（65岁以后）由于脑血管病、老年痴呆和神经系统退行性病变增多，癫痫发病率又见上升。

一旦出现癫痫症状，一定要及时到正规专业的癫痫病医院，进行全面科学的诊断，患者一定要积极配合医生进行综合治疗，同时做好日常生活保健，积极做到养治结合，才能够尽快治愈疾病，争取早日康复。癫痫的诊断过程非常复杂，需要经过以下步骤：

1) 脑电图检查对癫痫和癫痫发作的分类有帮助，是诊断癫痫和癫痫发作的最重要的手段。临床怀疑癫痫的病例均应进行脑电图检查。需要注意的是，一般常规脑电图的异常率很低，约为10~30%。而规范化脑电图，由于其适当延长描图时间，保证各种诱发试验，特别是睡眠诱发，必要时加作蝶骨电极描记，因此明显提高了癫痫放电的检出率，可使阳性率提高至80%左右，并使癫痫诊断的准确率明显提高。

2) 如为继发性癫痫应进一步行头颅CT、头颅MRI、MRA、DSA等检查可发现相应的病灶。主要依据详细的病史资料、规范化的脑电图检查，必要时行录像脑电图检测等进行判断。

3) 腰穿CSF检查，可有改变。

癫痫的确诊主要有以下依据^[46]：

1) 确定癫痫病的病因。在对患者进行癫痫病的诊断时，首先要进行癫痫病和癫痫病综合征的分类，还需要确定症状性癫痫病患者的病因。脑血管病、皮质发育障碍、全身系统性疾病、肿瘤、寄生虫、脑外伤、药物、遗传代谢病以及中枢神经系统感染等都是常见的癫痫病因。查体中有无神经系统体征、全身性疾病等。然后选择有关检查，如头颅磁共振(MRI)、CT、血糖、血钙、脑脊液检查等，以进一步查明病因。

2) 病史资料。一份详细而又准确的病史是诊断癫痫的重要依据，在病史中应询问有无家族史、出生及生长发育情况、有无脑炎、脑膜炎、脑外伤等病史。当患者不能诉说发作过程时，需向目睹者仔细了解发作全过程，包括当时环境，发作时的姿态、面色、声音、有无肢体抽搐和其大致顺序。了解发作时有无意识丧失等，对诊断全面强直一阵挛发作非常关键。

3) 确定类型。癫痫疾病诊断还要确定其发作类型属于哪种，对于不同的发作类型，其治疗措施和预后均不相同，应根据患者病史、病发年龄、脑电检查结果和仔细分析其它资料来确定发作类型。

4) 实验室检测。主要有以下几种检测方式：

血象检测：血液化验的整个结果，包括红细胞数目、白细胞数目和血小板数目等。检查

可明白抗癫痫药物能否招致粒细胞减少，可逆性血小板减少，以至能否引起再生障碍性贫血和中毒性肝炎等疾病；

动态脑电图检查：能够连续 24 小时记载患者脑电波形，可捕捉每次异常信号，医生能够回放、研讨患者每一秒脑电细节，使误诊率大大降低，完成正确诊断，明白分型，系统治疗的准绳。从而到达临床治愈；

血药浓度检测：泛指药物在全血中的浓度。普通药物作用强度与药物在血浆中浓度成正比。检查可明白药物在患者体内代谢状况，可作为调整药物及剂量根据，判别患者药；

脑 CT 检查：构造影像学检查，适用于大脑器质性损伤招致癫痫发作者，检查可显现“癫痫灶”必要时分离核磁共振做更进一步详细检查。

值得注意的是，在癫痫的诊断过程中应该区别痫性发作还是非痫性发作，患者出现了发作性症状后，最重要的问题不是如何治疗，而是首先明确癫痫病的诊断。有资料显示，在难治性癫痫患者中，少部分患者根本不是癫痫病，而是非痫性发作，按照癫痫病治疗自然没有效果。有许多非痫性发作同样伴有意识、感觉、运动和精神的改变，必须与痫性发作区别，如晕厥、癔病性发作、短暂性脑缺血发作、低血糖、低血钙、梦游、精神障碍性疾病、运动障碍性疾病等。

表 4.2 不同的年龄组常见非痫性发作

年龄组	非痫性发作
新生儿	周期性呼吸、非惊厥性呼吸暂停、颤动
婴幼儿	屏气发作、非癫痫性强直发作、情感性交叉擦腿动作、过度惊吓症
儿童	睡眠肌阵挛、夜惊、梦魇及梦游症、发作性睡病、多发性抽动症、发作性运动诱发性运动障碍
成人	晕厥、癔病、短暂性脑缺血发作、偏头痛、精神病性发作

五、预防预后注意事项

预防癫痫病发生应注意以下几方面：[47-49]：

①优生优育，禁止近亲结婚。孕期头三个月，一定要远离辐射，避免病毒和细菌感染。规律孕检，分娩时避免胎儿缺氧、窒息、产伤等

②小儿发热时应及时就诊，避免孩子发生高热惊厥，损伤脑组织。还应看护好孩子，避免其发生头外伤。

③青年人、中年人、老年人应注意保证健康的生活方式，以减少患脑炎、脑膜炎、脑血

管病等疾病发生。

约 70% 癫痫患者的发作是可以得到控制的。经过 2~5 年正规的抗癫痫药物治疗，癫痫患者中 50%~60% 是可以恢复正常，患者可以和正常人一样地工作和生活。手术治疗和神经调控治疗可使部分药物难治性癫痫患者的发作得到控制或治愈，从一定程度上改善了难治性癫痫的预后。预防癫痫发作复发，应主要注意以下几方面：

① 生活规律，按时休息，保证充足睡眠，避免熬夜、疲劳等。避免长时间看电视、打游戏机等。加强体质锻炼，起居有规律，忌烟酒等刺激食物。

② 饮食清淡，多食蔬菜水果，避免咖啡、可乐、辛辣等兴奋性饮料及食物，戒烟、戒酒。避免服用含有咖啡因、麻黄碱的药物。青霉素类或沙星类药物有时也可诱发发作。

③ 按时、规律服药，定期门诊随诊。癫痫患者不能骤减或停服抗病药，以免引起癫痫持续状态。

④ 禁止驾驶汽车；禁止在海边或江河里游泳；不宜在高空作业、不操作机器等。如有发作预兆，应立即卧倒，避免跌伤。

4.1.2 癫痫异常脑电信号研究现状

在癫痫的辅助诊断方法中，到目前为止，最方便、最重要的和最有价值的手段之一是脑电图检查。每个神经细胞都进行着有节律的自发放电活动^[41]，脑电活动的基本节律是由 8~13 周/秒的 α 波所组成，但其相当恒定，且频率较低（10~20 周/秒，与神经细胞膜的周期性去极化相一致）。虽然大脑约有 100 亿个神经细胞^[41]，但每个神经细胞的电位差却非常小，这是因为整个脑的仅达百万分之一伏特，电位活动甚小。同时，神经细胞放电的总和在时间上是非同步性的，其主要原因是神经细胞的放电是分散的。若放电的时间集中（同步），则波幅高、持续时间短、波形尖。如果放电的时间越分散，那么就会使得总放电的外形的波幅越低，相应的波形越钝、持续时间也越长^[41]。

视频脑电图的意义：一般而言，异常脑电图最容易在睡眠浅期出现，所以睡眠脑电图是发现异常脑电波最好的方法之一。不使用安眠药睡眠最好。因为用安眠药入睡，很快进入深睡状态，错过最易发现异常脑电图的时机，而且研究表明，安眠药干扰脑电图结果，减低睡眠时异常脑电图出现的机会。所以，我们主张不用安眠药作脑电图，而应自然睡眠或剥夺睡眠作脑电图。剥夺睡眠就是让患者早晨或晚上一段时间不睡觉，这样，作检查时就很易入睡。

癫痫病人的脑电图变化有 2 类：

1) 一类是特异性异常，即对癫痫有诊断意义。

棘波：皮形像针尖一样尖锐，多为负相波，也可正相、双相或三相波。

尖波：典型尖波有急速上二升支和较缓慢下降支组成，呈锯齿状。

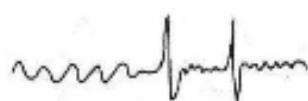
棘一慢综合波，波形成，均为负相波；典型 3 周/秒棘一慢波，为失神发作特殊波形。

尖一慢综合波：由尖波和慢波组成。

多棘一慢综合波：由几个棘波和一个慢波组成。

峰节律紊乱：是一种独特波形，即多棘一慢综合波^{41]}或高波幅的棘波尖波多棘波^[及慢波在部位上、时间上、幅度上杂乱且毫无规律地出现，多见婴儿痉挛症，预示有严重的脑损伤。]

常见癫痫波形如图 4.1 所示：



(a) 棘波



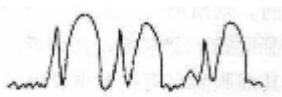
(b) 尖波



(c) 棘-慢波



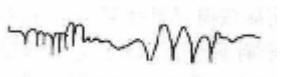
(d) 多棘-慢波



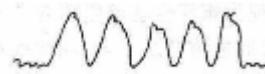
(e) 尖-慢波



(f) 高峰节律紊乱



(g) 14 或 6 周/秒正相棘波



(h) 发作性节律波

图 4.1 常见癫痫波形图

2) 另一类是非特异性的异常，并不能确定为癫痫放电，但与癫痫或抗癫痫药物应用等有关，如弥漫性慢波、局限性慢波。癫痫发作时脑电图肯定异常放电，但有些癫痫病人在癫痫间歇期脑电图正常。有些癫痫病人癫痫间歇期可有异常放电活动，但这种异常放电也与癫痫发作一样具有特发性和一过性的特点。在脑电图描记过程中不一定遇上癫痫样放电。据统计，癫痫病人间歇期脑电图异常率约 80%，所以脑电图正常不能排除癫痫。那么脑电图异常是否可以确诊为癫痫？研究证明，少数正常人脑电图有时可出现典型的癫痫波，但从来没有癫痫发作。有人会说癫痫波的人暂时不发生癫痫，将来会发生癫痫，这是没有根据的。有人在正常儿童中进行调查，发现有千分之几的正常儿童脑电图有典型的癫痫波。经多年追踪观察，这些儿童发生癫痫的比例不比当初脑电图正常的儿童大。因此，单凭脑电图异常不能诊断癫痫，也不能预言将来必然会发生癫痫。

4.2 异常脑电信号不可逆性分析及意义

4.2.1 基于符号化相对熵的异常脑电信号不可逆性分析基本理论

熵的本质是变化的方向性和时间的方向性，熵指明了不可逆过程的方向性。熵值较小对应于比较有序的状态，熵值较大对应于比较无序的状态。可逆过程的熵值为零，现实中是不存在可逆过程的。过程的不可逆程度越大，熵产越大，即能量耗散越大^[50]。在熵增原理的前提下，物理学和统计力学紧密联系起来，实现了利用统计力学来量化研究生命现象。

本章在熵增原理的基础上，引入一种分析脑电信号时间不可逆特性的新方法——符号相对熵^[1]。相对熵量化了熵增和不可逆性之间的关系，将物理学和统计力学联系起来^[1]。本文提出了一种新的使用过程的前向概率和后向概率计算符号相对熵的方法，利用符号相对熵来估计能量耗散。并运用该方法分别对癫痫及正常脑电信号^[1]进行处理，最后针对实验结果进行了分析讨论。

时间序列的不可逆性和熵产之间的关系是物理学和统计力学的核心，两个随机过程间的相对熵定义如下：

$$D(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad (4.1)$$

相对熵是非负的, 当且仅当 $p(x)$ 和 $q(x)$ 相同时才为零。但是 $D(q \parallel p) \neq D(p \parallel q)$, 即相对熵不具有对称性^[1]。

熵产可用来衡量这两个随机过程间的可区别性。不同情况下的熵产和相对熵间的关系已经推导处理, 也经过验证。对于非平衡态, 每个时间单位的熵产为^[1]:

$$\frac{\langle s \rangle}{k} = \lim_{t \rightarrow \infty} \frac{1}{t} D[p(\{x(\tau)\}_{\tau=0}^t) \parallel p(\{x(t-\tau)\}_{\tau=0}^t)] \quad (4.2)$$

其中 k 是波尔兹曼常数, $p(\{x(\tau)\}_{\tau=0}^t)$ 是一个给定轨迹 $\{x(\tau)\}_{\tau=0}^t$ 的概率分布。对于静态轨迹, $\tilde{x}(\tau)$ 唯一确定熵产 s ^[1]。

为了克服噪声等因素的影响, 本文创新性的提出了使用符号相对熵算法来分析癫痫脑电信号不可逆性^[1]。该算法将原始序列符号化与相对熵分析法结合, 即先对原始脑电信号脑电 X 进行符号化及编码处理, 得到符号化序列 \dot{X} 。再计算其相对熵来获得脑电信号的不可逆信息。

原始序列的符号化定义如下^[1]:

$$s_i(\dot{x}_i) = \begin{cases} 0: & \mu_1 < x_i \leq (1+a)\mu_1 \quad or \quad (1+a)\mu_2 \leq x_i < \mu_2 \\ 1: & (1+a)\mu_1 < x_i < \infty \quad or \quad -\infty < x_i < (1+a)\mu_2 \\ 2: & (1-a)\mu_1 \leq x_i < \mu_1 \quad or \quad \mu_2 < x_i \leq (1-a)\mu_2 \\ 3: & (1-a)\mu_2 < x_i < (1-a)\mu_1 \end{cases} \quad (4.3)$$

其中 μ_1 是原始序列中大于等于零的取样信号的平均值^[1], μ_2 是小于零的平均值, 取 $a=0.05$ 。符号化主要的意义是对噪声的影响不敏感^[48], 得到的结论较为严格, 而其应用的关键在于如何划分符号区域, 使得处理后的信号不会丢失时间序列的动力学特性。

符号化的相对熵^[1]定义如下:

$$D_n(p_F \parallel p_B) \equiv \frac{1}{n} \sum_{l=1}^n p(\dot{x}_1, \dots, \dot{x}_n) \log \frac{p(\dot{x}_1, \dots, \dot{x}_l)}{p(\dot{x}_l, \dots, \dot{x}_n)} \quad (4.4)$$

我们对符号化序列逐点向前向后求其概率分布^[1], 其中 p_F 为序列每一点的向前概率分布, p_B 为序列每点的向后概率分布, $D_n(p_F \parallel p_B)$ 为长度单位为 n 的符号相对熵, 代表了序

列的不可逆程度。由生理机制驱动产生一个随机过程，那么该生理活动的平均熵产就定义为描述这个随机过程的时间前向概率分布和后向概率分布间的符号相对熵。符号相对熵反映了随机过程前向分布和后向分布的差异性^[1]。相对熵越大，差异性越大。序列的不可逆性也就越大。

推导可得熵产：

$$\frac{\langle \dot{S} \rangle}{k} \geq d(p_F \parallel p_B) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} D_n(p_F \parallel p_B) \quad (4.5)$$

这个公式揭示出物理机制和时间序列的统计学信息有密切联系。公式左侧是一个纯物理量，而右侧是仅与序列 x_1, \dots, x_n 有关的纯统计量。这种联系是 Landauer's 定律的泛化,Landauer's 定律将熵产和逻辑不可逆性联系起来^[1]。

4.2.2 算法流程

根据公式(4.4)，我们可以通过计算一个时间序列前向分布和后向分布间的相对熵^[1]，来得到产生该序列的生理机制的熵产信息，据此我们进而可以判断该生理过程是正常的还是病态的。这对于脑类疾病的诊断和资料有着重大意义。计算过程包括符号化、编码、逐点求序列前向概率及后向概率、计算符号相对熵和熵产生率，最后使用 spss 统计分析软件对计算结果进行了假设检验^[1]，验证了该算法的有效性。算法流程如图 4.2 所示：

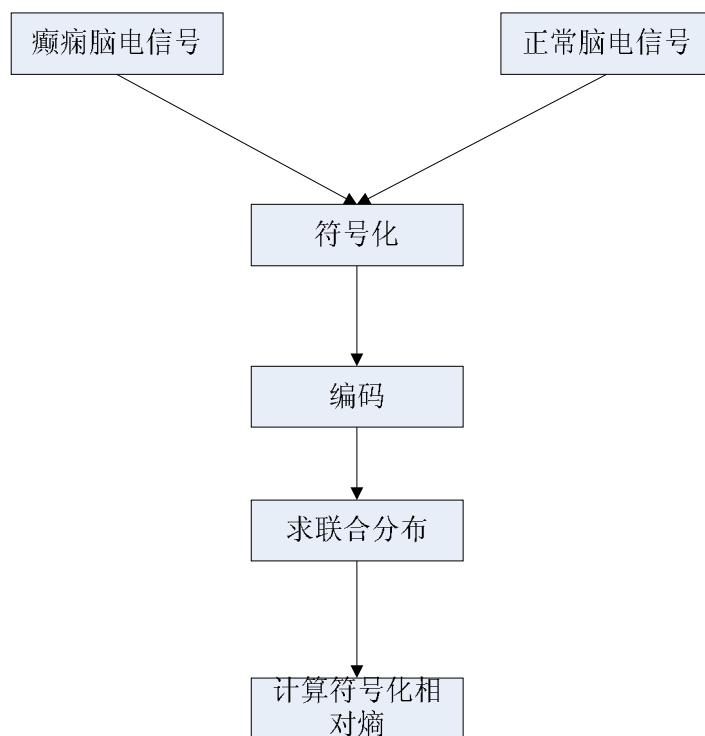


图 4.2 算法基本流程框图

4.3 异常脑电信号不可逆性分析实例仿真

4.3.1 实验数据

由于目前对于癫痫异常脑电信号的分析及研究较少，也缺乏权威准确的临床医学脑电信号数据库。本文采用了由南京军区总医院最新从临床诊断中采集到的一手资料。该数据库由癫痫患者脑电信号组和正常健康者脑电信号组构成。两组信号分别包括 22 个志愿者的脑电信号。采集设置是 24 导联，采样频率是 2400HZ。我们将这两组数据分别记为样本“DX”及样本“ZC”。

4.3.2 实验方法

根据公式(4.5)，我们可以通过计算一个时间序列前向分布和后向分布间的相对熵^[1]，来得到产生该序列的生理机制的熵产信息，据此我们进而可以判断该生理过程是正常的还是病态的。这对于脑类疾病的诊断和资料有着重大意义。计算过程包括符号化、编码、逐点求序列前向概率及后向概率、计算符号相对熵和熵产生率，最后使用 spss 统计分析软件对计算结果进行了假设检验，验证了该算法的有效性^[1]。对受训者样本“DX”和“ZC”脑电信号进行相对熵分析，首先对这两组非线性数据序列进行符号化，编码，相空间重构后，分别选择不同数据长度及编码长度来计算其相对熵，并进行显著性检测，然后进行多样本验证。

4.3.3 实验结果与分析

A.研究平均能量耗散与数据长度 L 的关系

在令编码长度 $N=6$ 的情况下，对受训者样本“DX”和“ZC”中的每组脑电信号序列分别取数据长度 $L=500, 1000, 1500, 2000, 2500, 3000$ 计算每个个体的能量耗散^[1]，并分别对癫痫样本和正常样本进行平均，相应的数据结果如表 4.3 所示：

表 4.3 (a) 不同数据长度下癫痫脑电信号的能量耗散

数据长度	L=500	L=1000	L=1500	L=2000	L=2500	L=3000
个体 1	19.81085	40.67052	43.36686	42.50156	57.41938	71.22415
个体 2	30.24416	54.83557	46.55905	52.58714	58.11723	59.7658
个体 3	20.33979	41.78387	54.98882	83.18252	72.1639	73.85071
个体 4	25.10452	33.68828	46.72515	49.13275	53.84513	61.77797
个体 5	21.97549	35.57275	42.25584	49.71311	58.89788	70.82998
个体 6	21.7799	35.86509	47.84856	52.78745	68.68949	69.28475
个体 7	17.28595	44.43482	51.48892	51.02646	93.99316	91.48337
个体 8	25.48921	38.64064	37.34898	49.72025	69.81185	69.23565
个体 9	22.81142	31.6176	50.20825	60.55734	48.08139	54.40507
个体 10	20.42973	40.10496	55.96165	70.72693	63.26988	69.41543
个体 11	19.41676	29.93515	34.56168	45.31198	53.24156	59.17338
个体 12	29.85834	33.73944	46.20366	71.11229	97.49033	100.458
个体 13	28.84793	44.52465	37.98498	53.86875	55.78626	63.7896
个体 14	30.29427	48.28469	51.70437	57.41252	55.7336	77.50818
个体 15	21.42983	28.41613	32.973	54.52377	59.25841	63.70834
个体 16	25.65726	39.39895	50.97584	46.08536	44.21489	53.20827
个体 17	28.87619	40.65662	59.85434	69.89293	84.3563	75.18366
个体 18	21.4119	32.2419	53.44605	62.55793	62.97229	68.25809
个体 19	31.67467	50.07897	66.92986	44.92772	62.76901	88.95077
个体 20	28.18716	37.11755	49.9387	51.02433	56.95421	61.14885
个体 21	29.50254	43.04195	65.55283	51.73263	74.76666	89.07693
个体 22	38.94031	47.54046	61.33805	59.76145	62.98063	64.50052

表 4.3 (b) 不同数据长度下的正常信号能量耗散

数据长度	L=500	L=1000	L=1500	L=2000	L=2500	L=3000
个体 1	29.19671	31.39137	59.2111	56.56946	57.49842	76.98026
个体 2	18.14057	46.47211	43.75264	34.30871	37.05838	47.24137
个体 3	32.44565	28.44075	46.80051	63.5101	78.35332	65.77942
个体 4	14.85113	26.24978	33.64257	55.47189	77.39902	79.40197
个体 5	24.62124	34.95382	39.05648	37.42131	47.09717	59.11977
个体 6	20.94655	29.2456	44.04989	43.10249	54.39667	73.61818
个体 7	26.79693	43.51281	42.40375	52.76486	54.39377	63.66467
个体 8	19.11594	29.14432	34.25659	62.23876	49.37638	68.85127
个体 9	35.28726	37.27091	54.81398	53.07294	64.83099	71.66678
个体 10	17.1253	32.95051	52.60411	48.06762	59.09735	72.91899
个体 11	31.28436	28.67107	54.69653	58.59354	52.78688	63.10899
个体 12	21.59362	32.69302	39.94697	59.58255	72.90037	72.71895
个体 13	24.42017	42.58728	36.00343	55.61833	60.85452	58.57103
个体 14	22.18523	39.9531	54.30385	70.29519	69.54569	63.59903
个体 15	22.57678	26.44601	46.64218	47.82201	63.99715	72.39143
个体 16	20.11523	29.37553	35.3498	49.49433	57.93485	59.67026
个体 17	23.68794	33.53115	44.29721	74.34505	74.29122	73.70252
个体 18	14.45981	30.48752	38.71638	48.18729	50.64246	56.27525
个体 19	20.96292	35.89376	27.09329	28.31007	37.12493	43.59335
个体 20	16.69378	19.72214	37.75747	56.72349	65.28854	87.91322
个体 21	18.5224	41.75767	48.73998	54.35403	56.50283	59.40732
个体 22	25.65909	42.23703	37.87331	59.81845	61.80556	92.13154

为更准确的对癫痫信号和正常信号的能量耗散进行对比发现，我们对每个样本中的不同个体进行平均，并取最大值最小值进行分析^[1]，结果如表 4.4 所示：

表 4.4 癫痫及正常脑电信号能量耗散与数据长度 L 的关系对比

数据长度	L=500	L=1000	L=1500	L=2000	L=2500	L=3000
癫痫平均值	25.42583	39.64503	49.46434	55.91578	64.3097	70.73807
正常平均值	22.75857	33.77215	43.27327	53.16693	59.23529	67.37843
癫痫最小值	17.28595	28.41613	32.973	42.50156	44.21489	53.20827
正常最小值	14.45981	19.72214	27.09329	28.31007	37.05838	43.59335
癫痫最大值	38.94031	54.83557	66.92986	83.18252	97.49033	100.458
正常最大值	35.28726	46.47211	59.2111	74.34505	78.35332	92.13154

得到平均能量耗散与数据长度 L 的关系^[1]直观对比如图 4.3 所示：

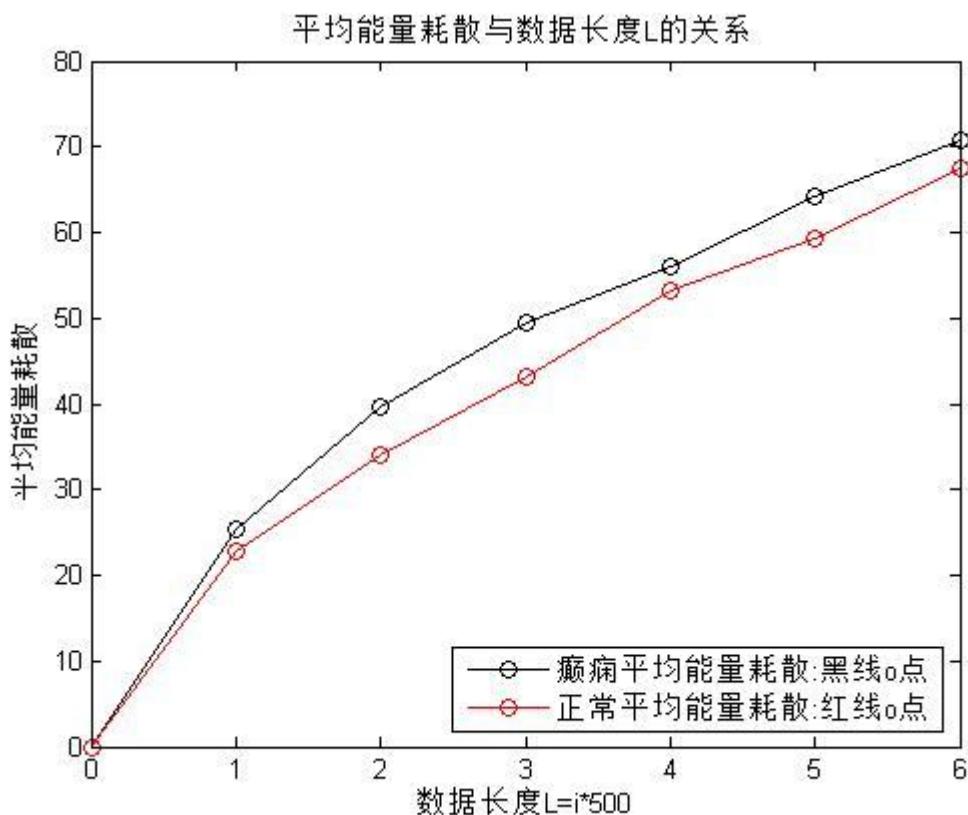


图 4.3 平均能量耗散与数据长度 L 的关系

分析图 4.4 癫痫与正常能量耗散曲线^[1]可知：

- 1) 当数据长度增大过程中，无论是正常还是癫痫脑电信号的平均能量耗散值也呈增大趋势。
- 2) 癫痫患者的平均能量耗散高于正常人的，在数据长度 $L=1000, 1500, 2500$ 时二者有相对较显著的差异。

从数据的精准性考虑，数据长度越大，其统计概率分布越接近实际分布，相应的准确性

越高^[1]。但从计算量以及计算速度来讲，数据长度越小，那么算法速度也越快，综合我们的实验结果，取数据长度 L=1000 时，即能兼顾处理速度，又可以保证实验精度^[1]。

B.研究平均能量耗散与编码长度 N 的关系

在令数据长度 L=1000 的情况下，对受训者样本“DX”和“ZC”中的每组脑电信号序列分别取编码长度 N=1, 2, ……, 10，计算每个个体的能量耗散^[1]，并分别对癫痫样本和正常样本进行平均，不同编码长度下癫痫信号及正常信号的能量耗散数据，如表 4.5 所示：

表 4.5 不同编码长度下癫痫信号及正常信号的能量耗散

编码长度	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
个体 1	14.13	35.40	38.37	40.24	38.63	37.29	36.50	36.60	36.45	37.43
个体 2	12.97	28.88	31.48	30.53	28.85	27.62	26.61	26.54	26.44	26.00
个体 3	13.25	31.24	41.14	42.10	43.88	44.45	45.00	46.18	47.47	47.94
个体 4	13.96	33.83	38.24	38.33	37.39	37.46	35.75	35.37	35.64	36.59
个体 5	13.13	33.68	37.79	36.77	36.00	34.31	33.00	31.45	29.97	28.33
个体 6	13.41	32.54	41.03	40.46	39.08	39.41	37.97	36.37	36.79	35.92
个体 7	13.06	34.33	40.75	42.82	42.37	43.07	42.33	42.43	44.25	45.37
个体 8	14.35	32.29	35.55	35.90	34.63	33.69	32.84	32.35	32.05	30.85
个体 9	12.29	35.81	48.77	47.66	46.08	43.24	42.43	41.38	40.51	40.13
个体 10	12.34	29.92	40.75	45.91	47.65	48.85	49.23	49.43	49.23	48.79
个体 11	13.07	30.05	34.83	33.11	31.22	30.71	30.37	29.35	28.15	27.94
个体 12	12.03	30.99	40.72	46.39	49.17	49.30	48.99	47.83	46.09	43.90
个体 13	13.52	35.78	41.86	41.99	42.49	43.55	44.12	44.03	43.77	43.55
个体 14	13.83	34.80	35.71	34.21	32.16	31.21	31.02	31.03	31.86	31.91
个体 15	13.79	38.25	46.34	50.61	52.94	53.16	50.79	49.04	47.33	46.24
个体 16	14.13	33.73	37.15	38.37	36.35	34.52	33.04	32.32	31.60	31.12
个体 17	14.03	36.15	46.24	46.61	46.03	44.97	44.24	43.68	43.34	43.52
个体 18	13.21	35.49	44.91	46.03	46.63	46.58	47.22	49.72	51.39	52.48
个体 19	12.16	32.86	44.45	47.88	45.90	44.10	42.41	40.44	38.84	37.37
个体 20	14.40	36.16	39.72	39.05	37.98	35.90	33.56	31.74	30.31	30.36
个体 21	13.28	29.79	37.15	39.47	39.63	39.42	39.11	38.91	38.57	38.77
个体 22	12.90	32.69	40.29	42.27	41.52	39.52	38.63	37.32	37.68	38.34

表 4.5 (b) 不同编码长度下正常信号的能量耗散

编码长度	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
个体 1	12.64	35.08	48.11	50.95	50.80	50.23	49.32	48.02	46.59	45.74
个体 2	16.01	35.67	39.85	40.59	40.55	42.20	43.72	45.12	46.59	47.86
个体 3	13.77	31.56	37.65	39.44	37.51	35.78	34.79	33.95	33.89	33.45
个体 4	14.54	26.48	25.39	23.54	24.28	26.67	30.08	32.68	34.10	35.11
个体 5	13.78	29.79	31.95	33.42	34.04	34.75	36.28	38.29	40.08	40.84
个体 6	14.07	28.78	30.61	28.59	27.95	28.14	30.03	32.20	34.53	37.52
个体 7	14.52	31.56	32.95	32.35	31.11	30.90	30.43	30.09	30.42	30.45
个体 8	13.38	34.31	44.62	44.50	43.69	42.10	40.33	38.27	36.75	35.49
个体 9	13.35	31.44	37.22	37.72	36.87	36.11	35.56	34.58	33.82	32.54
个体 10	11.84	33.46	49.97	51.06	48.67	48.27	47.09	46.67	45.88	45.17
个体 11	14.13	32.72	35.65	36.32	36.17	34.95	34.95	34.23	33.74	33.06
个体 12	12.95	32.87	38.25	39.51	39.06	36.33	34.18	32.40	31.15	30.94
个体 13	13.00	35.78	33.62	31.24	28.74	26.19	24.28	23.06	23.58	24.12
个体 14	14.42	35.68	43.22	45.28	44.86	43.81	42.96	41.72	40.46	39.37
个体 15	14.22	34.78	37.26	38.16	37.75	37.76	38.09	37.95	39.38	41.30
个体 16	12.53	31.63	35.79	37.39	36.77	34.28	32.39	32.57	33.42	33.67
个体 17	13.98	34.12	39.69	41.38	41.20	40.64	41.51	43.07	44.14	44.65
个体 18	12.66	35.63	39.97	40.41	39.00	36.35	33.05	30.04	27.56	25.89
个体 19	14.13	27.73	27.10	26.69	25.65	25.26	27.02	29.20	30.59	31.94
个体 20	13.57	34.03	37.13	37.05	35.64	33.80	32.85	31.48	30.94	29.94
个体 21	13.03	35.02	44.21	43.97	42.12	40.02	40.18	39.50	38.36	37.99
个体 22	14.81	32.50	33.16	33.13	32.82	31.82	33.17	36.16	38.38	39.95

对癫痫信号和正常信号的能量耗散进行取平均值、最大值、最小值进行对比分析^[1],结果如表 4.6 所示:

表 4.6 癫痫信号和正常信号的能量耗散与编码长度 N 的关系对比

编码长度	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
癫痫均值	13.33	33.39	40.15	41.21	40.75	40.11	39.33	38.80	38.53	38.31
正常均值	13.70	32.76	37.43	37.85	37.06	36.20	36.01	35.97	36.11	36.23
癫痫最小值	12.03	28.88	31.48	30.53	28.85	27.62	26.61	26.54	26.44	26.00
正常最小值	11.84	26.48	25.39	23.54	24.28	25.26	24.28	23.06	23.58	24.12
癫痫最大值	14.40	38.25	48.77	50.61	52.94	53.16	50.79	49.72	51.39	52.48
正常最大值	16.01	35.78	49.97	51.06	50.80	50.23	49.32	48.02	46.59	47.86

得到平均能量耗散与编码长度 N 的关系如图 4.4 所示：

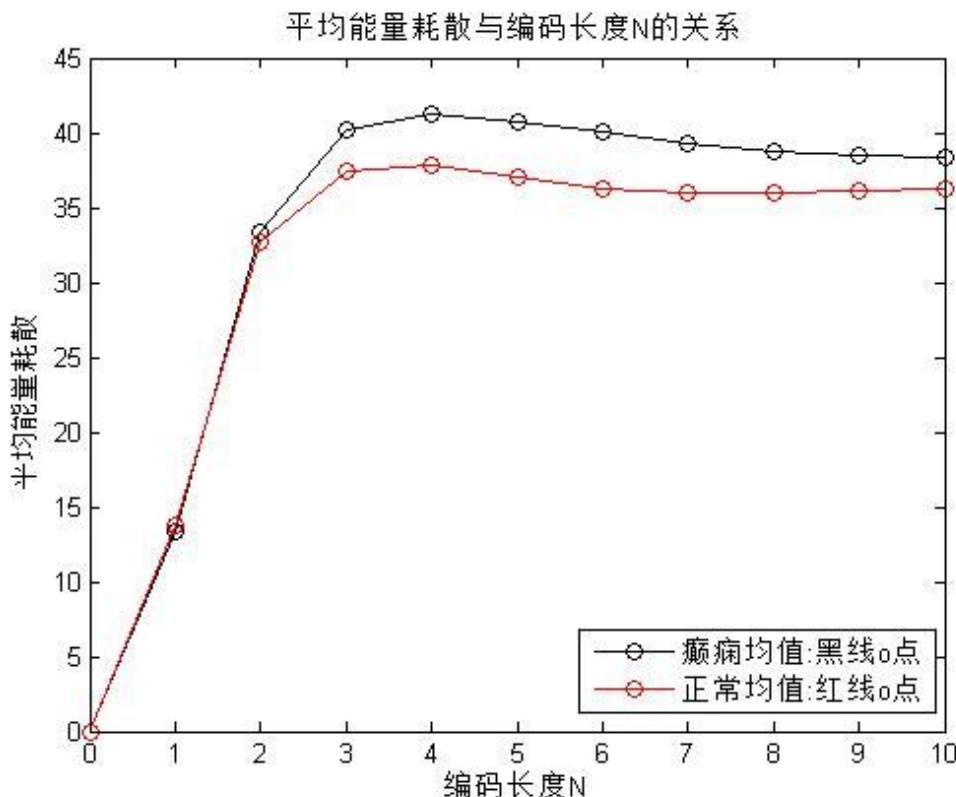


图 4.4 平均能量耗散与编码长度 N 的关系

分析图 4.4 平均能量耗散与编码长度 N 的关系可知^[1]:

1) 在编码长度 N 从 1 到 3 的递增过程中，癫痫和正常信号的平均能量耗散都成迅速增大趋势。从 N 取 3 之后平均能量耗散缓慢减小。

2) 癫痫患者的平均能量耗散明显高于正常人的，且在 N=4, 5 及 6 时差异最显著。

考虑到实验效果的明显性，n 应该取大些^[1]；考虑到算法的复杂度会影响实验处理速度以及在临床应用上的实时性，n 应该取小一些。综合考虑上述因素，n 取 6。

C.研究在数据长度 L=1000,编码长度 N=6 条件下癫痫患者与正常人的能量耗散的差异

对受训者样本“DX”和“ZC”中的每组脑电信号序列分别取编码长度 N=6 及数据长度 L=1000，计算每个个体的能量耗散，并分别对青年样本和老年样本进行平均^[1]，得到平均能量耗散与健康状况的关系如图 4.5 所示：

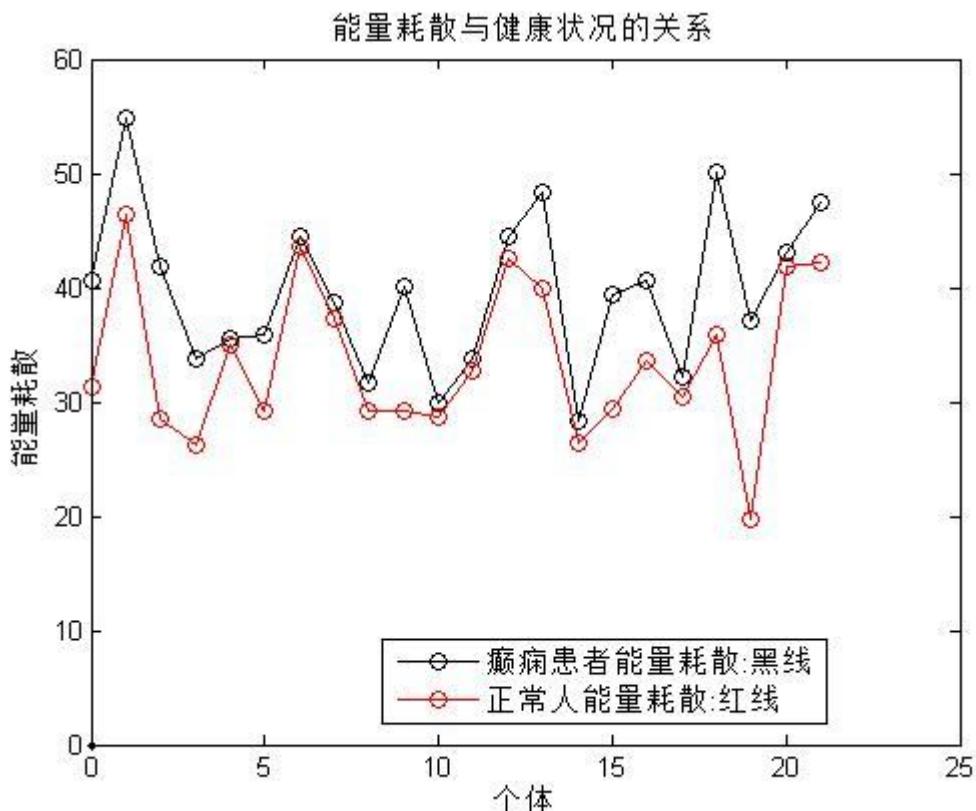


图 4.5 平均能量耗散与健康状况的关系

分析图 4.5 平均能量耗散与健康状况的关系可知：癫痫患者的能量耗散明显高于正常人的，个体间的能量耗散差异较大^[1]。

D.统计分析与假设检验

为进一步验证本文计算结果的准确性及算法的有效性，使用 spss 统计分析软件对计算结果进行了假设检验^[1]。主要方法如下：

对癫痫患者和正常人的能量耗散值的差异显著性进行假设检验，假设二者没有显著差异。

使用 spss 对两组样本数据 DX、ZC 进行独立样本 T 检验^[1]，结果如表 4.7 所示：

表 4.7 青老年平均能量耗散的独立样本 T 检测

组统计量				
V2	N	均值	标准差	均值的标准误
癫痫	22	33.772149	6.7870292	1.4469995
正常	22	39.645026	6.8662652	1.4638927

独立样本检验									
样本		方差方程的 Levene 检验		均值方程的 t 检验					
		F	Sig.	t	df	Sig.(双侧)	均值差值	标准误差	差分的 90% 的置信区间
癫痫	假设方差相等	0.004	0.949	2.853	42	0.007	5.873	2.058	1.719 10.027
正常	假设方差不相等	0.004	0.949	2.853	41.994	0.007	5.873	2.058	1.719 10.027

根据表 4.7 分析癫痫患者及正常人能量耗散差异的显著性，首先假设二者没有差异,分析表 4.7 可以得到以下结论^[1]:

①由表 1 可知，在假设方差相等的情况下，自由度(df)为 42，查找 t 值表可知理论 t 值 $t(df)_{0.05}=t(42)=2.021$ ，而样本 t 值为 2.853，大于 $t(42)$ ，且 $Sig(\text{双侧})=0.007<0.05$ ，假设不成立，所以癫痫患者和正常人脑电信号能量耗散水平差异显著。

②由表 1 可知，在假设方差不相等的情况下，自由度(df)为 42，查找 t 值表可知理论 t 值 $t(df)_{0.05}=t(42)=2.021$ ，而样本 t 值为 2.853，大于 $t(42)$ ，且 $Sig(\text{双侧})=0.007<0.05$ ，假设不成立。所以癫痫患者和正常人脑电信号能量耗散水平差异显著，证明该算法可以有效的对癫痫患者和正常人的脑电信号进行区分^[1]。

由前言中讨论可知，在一个孤立系统内部自然发生的过程总使系统不可逆的趋于熵取极大值的平衡态^[1]。将人的生命视作一个过程时,随着年龄的增长，生命的不可逆程度越大，熵产越大，即能量耗散越大^[1]。熵值较小对应于比较有序的状态，也就是脑电活动的健康状态；熵值较大对应于比较无序的状态，也就是脑电活动的消极状态。这与我们的实验结果也是一致的，证明使用相对熵来分析脑电信号的不可逆性是可行的，我们可以以平均能量耗散作为特征值来衡量脑电信号是否处于健康状态^[1]。

4.4 小结

本文首先介绍了癫痫脑电信号研究现状^[1]。接着介绍基于符号相对熵的脑电信号时间不可逆性研究方法。根据本文提出的算法，我们首先对脑电信号进行相应的符号化及编码处理后，再计算过程包括符号化、编码、逐点求序列前向概率及后向概率、计算符号相对熵和熵产生率，最后使用 spss 统计分析软件对计算结果进行了假设检验，验证了该算法的有效性^[1]。主要的工作内容如下：

(1)基于过程不可逆特性提出了一种新的符号相对熵的计算方法，即使用过程的前向概率和后向概率来计算相对熵，从而获得过程的不可逆程度信息^[1]。

(2)基于该算法对癫痫及正常脑电信号进行了分析,结果表明,癫痫患者脑电信号的能量耗散明显高于正常人的,表明能量耗散可以作为衡量一个生物过程是否处于健康积极状态的参数^[1]。

第五章 基于符号化部分互信息的多参量生物电信号分析

5.1 信息熵研究现状

在传统热力学中，熵被定义为对系统的宏观测定，并没有涉及概率分布，而概率分布是信息熵的核心定义。物理学家和化学家对一个系统自发地从初始状态向前演进过程中，遵循热力学第二定律^[52-54]而发生的熵的变化更感兴趣。

克劳德·艾尔伍德·香农首次将热力学的熵引入到信息论中^[55]。在信息论中，熵用来测量不确定性的大小，熵越高，意味着传输的信息越多，反之则相反。熵代表了在被接收之前，信号传输过程中损失的信息量，又称为信源熵、平均自信息量。根据 Jaynes 的观点^[56-59]，热力学熵可以被视为是香农信息理论的一个应用：热力学熵被定义为与要进一步确定系统的微观状态所需要的更多香农信息的量成正比^[55]。

一个取值范围为 $\{x_1, \dots, x_n\}$ 的随机变量 X 的香农熵^[60-61]定义如方程 5.1 所示：

$$H(X) = E(I(X)) \quad (5.1)$$

其中， E 表示期望函数，而 $I(X)$ 是 X 的信息量。熵是为了刻画一个随机过程的不确定性、混乱程度。当一个随机过程是均匀分布时，熵值最大。这是由于均匀分布时，限定最小，相应的不确定性就越大，熵取最大值。

对于离散型随机变量 X ，信息熵^[55]可以表示为方程 5.2 所示：

$$H(X) = \sum_{i=1}^n p(x_i)I(x_i) = -\sum_{i=1}^n p(x_i)\log_b p(x_i) \quad (5.2)$$

这里 b 是对数所使用的底，通常是 2，自然数 e ，或者 10。

熵具有以下两个基本特征：

- (1) 熵均大于等于零，即 $H(X) \geq 0$ ；
- (2) 若随机过程 S 的随机总数是 N ，那么 S 的信息熵 $H_s \leq \log_2 N$ 。当且仅当各事件等概率时等号成立，此时随机过程 S 的信息熵最大。

5.1.1 互信息基本原理

互信息^[62-69]用于度量两个随机变量间的相互性。一般来讲，信道中总是存在着噪声和干

扰。信源发出信息 X ，在接收端收到的信息 Y 已经受到干扰影响。先验概率 $p(x)$ 用于描述信源发出 X 的概率，后验概率 $p(x|y)$ 用于描述信宿收到 Y 后推测信源发出 X 的概率。那么 Y 对 X 的互信息量定义为 X 的后验概率与先验概率比值的对数。

两个随机变量 X 和 Y 的平均互信息^[70-75]定义为：

$$I(X, Y) = H(X) - H(X | Y) \quad (5.3)$$

$I(X, Y)$ 反映了随机变量 X 和 Y 间的相关性^[55]，表示在获得其中一个变量的知识时，对另一个变量的不确定性减小的量。其中联合熵 $H(X, Y)$ 由 XY 的联合分布 $\{p_{xy}\}$ 计算得到。

平均互信息熵^[76-81]具有以下基本特征：

(1) 非负性,即

$$I(X, Y) \geq 0 \quad (5.4)$$

当 X 和 Y 统计独立时等号成立。即：通过一个信道获得的平均信息量不可能是负的，一般总能获得一些信息，只有在 X 和 Y 统计独立的极端情况下，才接收不到任何信息，即 $I(X, Y) = 0$ 。

(2) 极值性,即

$$I(X, Y) \leq H(X) \quad (5.5)$$

这一性质的直观含义为：接受者通过信道获得的信息量不可能超过信源本身固有的信息量。只有当信道为无损信道，即信道疑义度 $H(X | Y) = 0$ 时，才能获得信源中的全部信息量。

(3) 对称性,即

$$I(X, Y) = I(Y, X) \quad (5.6)$$

$I(X, Y)$ 表示接收 Y 后获得的关于 X 的信息量，而 $I(Y, X)$ 表示发出 X 以后达到的关于 Y 的信息量。当输入 X 与输出 Y 统计独立时有：

$$I(X, Y) = I(Y, X) = 0 \quad (5.7)$$

即此时不能由一个随机变量获得关于另一个随机变量的信息。当输入 X 与输出 Y 一一对应时，则有：

$$I(X, Y) = I(Y, X) = H(X) = H(Y) \quad (5.8)$$

即从一个随机变量可以获得另一个随机变量的全部信息。

(3) 平均互信息与其他熵的关系

$$I(X, Y) = H(X) - H(X | Y) = H(Y) - H(Y | X) \quad (5.9)$$

$$I(X, Y) = H(X) + H(Y) - H(XY) \quad (5.10)$$

其中 $H(X | Y)$ 和 $H(Y | X)$ 是条件熵, $H(X | Y)$ 是信道疑义度, 表示信道符号通过有噪声信道传输后平均损失的信息量, $H(Y | X)$ 为噪声熵(或散布度), 是由信道中的噪声引起的, 表示在已知 X 的条件下, 对随机变量 Y 存在的平均不确定性。 $H(X, Y)$ 是输入 X 与输出 Y 的联合熵,。当随机变量 X 和 Y 相互独立时, 互信息量为零^[82-88]。它们之间的关系可由图 5-1 表示:

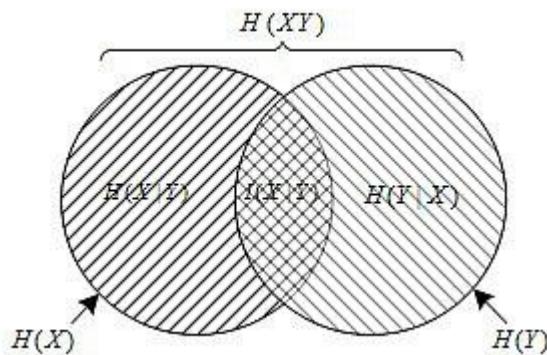


图 5.1 $I(X, Y)$ 与各类熵的关系图

(4) 平均互信息具有凸函数特性

推理可得, $I(X, Y)$ 完全是信源 X 的概率分布 $P(x)$ 及信道转移概率 $P(y | x)$ 的函数, 即

$$I(X, Y) = \sum_{xy} P(x)P(y | x) \log \frac{P(y | x)}{\sum_{y|x} P(y | x)} = f[P(x), P(y | x)] \quad (5.11)$$

$I(X, Y)$ 是信源概率分布 $P(x)$ 的 \cap 型凸函数, 其与信源分布 $P(x)$ 的关系曲线如图 5.2 所示:

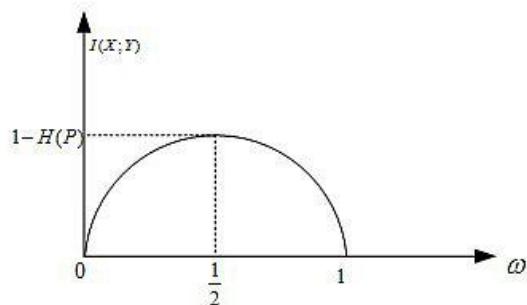


图 5.2 $I(X, Y)$ 与信源分布 $P(x)$ 的关系曲线图

$I(X, Y)$ 是信源概率分布 $P(y|x)$ 的 U型凸函数, 其与信源分布 $P(y|x)$ 的关系曲线如图 5-3 所示:

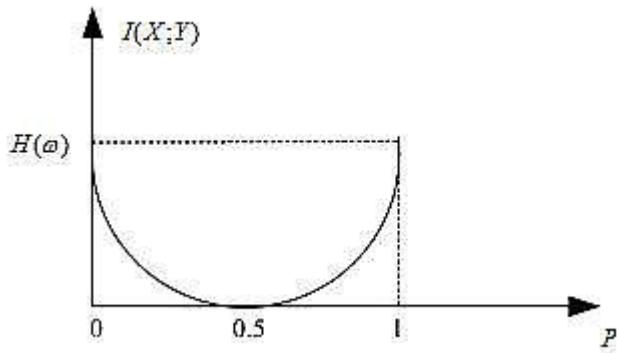


图 5.3 $I(X, Y)$ 与信源分布 $P(y|x)$ 的关系曲线图

5.1.2 部分互信息基本原理

部分互信息 $I(X, Y|Z)$ 是 XY 相交但不包括 Z 的部分, 其定义式如下:

$$I(X, Y|Z) = H(X, Z) + H(Y, Z) - H(Z) - H(X, Y, Z) \quad (5.12)$$

$I(X, Y|Z)$ 表示在已知 Z 的情况下 Y 提供的关于 X 的平均信息量^[55], 其物理意义如图 5.4 所示:

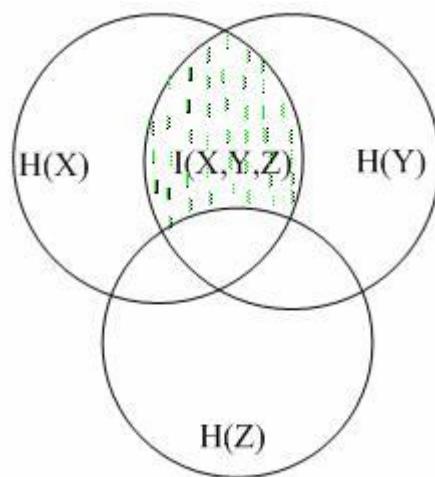


图 5.4 部分互信息 $I(X, Y|Z)$ (间断条纹部分)

对于相同的 Z , 互信息 PMI 是对称的^[55], 即 $I(X, Y|Z) = I(Y, X|Z)$, 同样有 $0 \leq I(X, Y|Z)$, 当且仅当 $X Y$ 相互独立时取零, 特别是当 X 或 Y 是 Z 的函数时。随机变量间的互信息越大, 其耦合程度越大。

5.2 时间序列耦合分析

5.2.1 基于符号化部分互信息算法的时间序列耦合分析基本理论

为了克服噪声等因素的影响，本文创新性的提出了使用符号化的部分互信息来对生物电信号进行耦合分析^[55]。原始序列的符号化定义如下：

$$s_i(\dot{x}_i) = \begin{cases} 0: & \mu_1 < x_i \leq (1+a)\mu_1 \quad or \quad (1+a)\mu_2 \leq x_i < \mu_2 \\ 1: & (1+a)\mu_1 < x_i < \infty \quad or \quad -\infty < x_i < (1+a)\mu_2 \\ 2: & (1-a)\mu_1 \leq x_i < \mu_1 \quad or \quad \mu_2 < x_i \leq (1-a)\mu_2 \\ 3: & (1-a)\mu_2 < x_i < (1-a)\mu_1 \end{cases} \quad (5.13)$$

其中 μ_1 是原始序列中大于等于零的取样信号的平均值^[55]， μ_2 是小于零的平均值，取 $a=0.05$ 。符号化主要的意义是对噪声的影响不敏感^[55]，得到的结论较为严格，而其应用的关键在于如何划分符号区域，使得处理后的信号不会丢失时间序列的动力学特性。当 a 取 0.04 到 0.07 时都可得到原信号的大尺度信息，如果 a 取小于 0.04 或大于 0.07 时，会得不到较为合理的统计特性。这是因为在符号化的过程中，如果 a 的值过大或过小，会导致细节信息的丢失。故取 $a=0.05$ ，既能去掉原信号噪声影响，又能较好的捕捉信号中的动态信息^[55]。

为了取得更好的分析效果，本文将原始序列符号化与部分互信息分析法结合，即先对多参数的生物电原始信号脑电 X 、心电 Y 、肌电 Z 进行符号化及编码处理，得到符号化序列 \dot{X} 、 \dot{Y} 、 \dot{Z} ，再计算其部分互信息来获得多参量生物电信号的耦合信息^[55]，符号化的部分互信息定义如下：

$$I(\dot{X}, \dot{Y} | \dot{Z}) = H(\dot{X}, \dot{Z}) + H(\dot{Y}, \dot{Z}) - H(\dot{Z}) - H(\dot{X}, \dot{Y}, \dot{Z}) \quad (5.14)$$

5.2.2 算法流程

本章主要算法流程如图 5.5 所示^[55]：

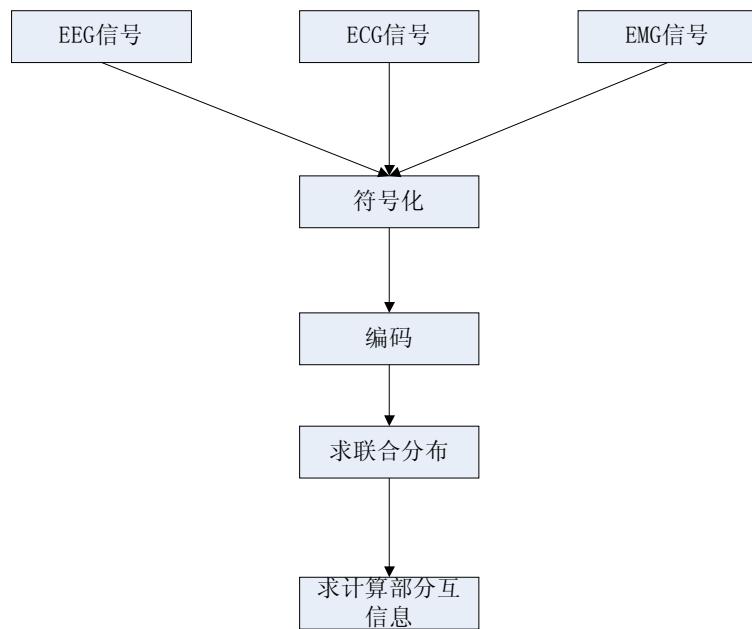


图 5.5 符号化部分互信息算法流程

5.3 多参量生物电信号耦合分析实例仿真

5.3.1 实验数据

本文使用的睡眠数据来自PhysioBank的MIT-BIH Polysomnographic Database^[55]。该库中的记录是多参数睡眠数据，包括1导EEG(electroencephalogram)信号，1导ECG(electrocardiosignal)信号，1导EOG(eyectro-oculogram)信号，1导EMG(electromyographic signal)信号等多导睡眠信号，记录长度为6个小时，数据采样频率250Hz，每份记录的数据都附带着以30s为一个分期的注释信息，本文的实验结论判定依此注释为参考。

本文所用数据，分别采用了15组受试者睡眠和清醒期的多参数生物电信号数据中的1导EEG(C3-O1)脑电信号、1导ECG心电信号、1导EMG肌电信号,提取其中的清醒期和NREM睡眠I期的若干组信号^[55]，分别记为样本Sleep和Weak。

5.3.2 实验方法

根据本文提出的算法，我们首先根据公式(5.13)对生物电原始信号各参量 EEG、ECG、EMG 进行相应的符号化及编码处理^[55]，再根据公式(5.14)计算生物电信号样本 Sleep 和 Weak 各参量间的部分互信息 $I(EEG,ECG|EMG)$ ，从而分别获得睡眠期和清醒期生物电信号的耦合信息。最后使用 spss 统计分析软件对计算结果进行了假设检验^[55]，验证了该算法的有效性。

以生物电信号的耦合信息作为参数，我们可以判断该生物机制是处在活跃的还是消极的状态，这对于一些身体内部器官健康状况的诊断具有重要意义^[55]。

5.3.3 实验结果及分析

A. 研究多参量生物电信号耦合程度与数据长度 L 的关系

对受训者样本 15 组 Sleep 和 15 组 Wake 中的每组生物电信号序列分别取数据长度 $L=500, 1000, 1500, 2000, 2500, 3000^{[1]}$ ，并对每个时间序列进行符号化及编码处理^[55]，分别计算每个个体的部分互信息，并分别对睡眠期部分互信息和清醒期部分互信息进行平均，得到睡眠期和清醒期耦合程度与数据长度 L 的关系如表 5.1 所示^[55]：

表 5.1 (a) 睡眠期耦合程度与数据长度 L 的关系

数据长度	L=500	L=1000	L=1500	L=2000	L=2500	L=3000
个体 1	0.746467	0.911585	1.022546	1.074904	1.006912	1.617236
个体 2	1.085356	1.218242	1.103188	1.25241	0.918898	1.066885
个体 3	0.714432	0.904197	0.92211	0.921369	0.961439	0.954379
个体 4	0.90191	0.994825	1.345677	1.397659	1.585234	1.418066
个体 5	0.607273	1.161577	1.032049	1.366061	1.462944	1.218475
个体 6	1.018142	0.606597	0.894518	1.220364	1.288235	1.330156
个体 7	1.308735	1.24784	0.905002	1.008851	1.128547	1.40229
个体 8	0.876621	1.440305	1.515802	1.642767	1.403927	1.595264
个体 9	0.738263	0.983544	1.108616	1.084664	0.820673	0.986426
个体 10	2.986742	2.582212	2.540615	2.420075	2.163659	2.47706
个体 11	2.99702	2.459386	2.226721	2.242771	2.184947	1.834184
个体 12	2.332917	2.633299	2.620442	2.395332	2.097299	1.999684
个体 13	2.632574	2.639486	2.370176	2.199001	1.65508	1.650283
个体 14	2.232662	2.060926	1.649337	1.692031	1.558563	1.575421
个体 15	1.308735	1.440305	1.515802	1.642767	1.403927	1.575421

表 5.1 (b) 清醒期耦合程度与数据长度 L 的关系

数据长度	L=500	L=1000	L=1500	L=2000	L=2500	L=3000
个体 1	1. 239901	0. 773337	1. 275136	1. 583887	1. 709895	1. 777105
个体 2	0. 852019	0. 916577	1. 25847	1. 605325	1. 714168	1. 687208
个体 3	1. 483256	1. 410721	0. 952402	0. 865435	1. 031482	1. 257231
个体 4	1. 310437	1. 427569	1. 331876	1. 641146	1. 558212	1. 525436
个体 5	1. 417751	1. 385671	1. 40453	1. 579076	1. 554442	1. 562558
个体 6	1. 482132	1. 699755	1. 841494	1. 752913	1. 87462	1. 871387
个体 7	1. 176077	1. 45364	1. 927175	2. 01316	2. 013998	2. 049036
个体 8	1. 20086	1. 403722	1. 541157	1. 378325	1. 50795	1. 683672
个体 9	1. 157241	1. 384463	1. 833409	1. 81987	1. 48883	1. 657214
个体 10	2. 566605	2. 852547	3. 18648	3. 160223	3. 090469	3. 497504
个体 11	2. 362226	2. 994935	3. 177909	3. 138265	3. 071831	3. 196351
个体 12	2. 956803	3. 171682	3. 428297	3. 426001	3. 613892	3. 528177
个体 13	2. 33059	2. 991234	2. 917144	2. 513066	2. 457104	2. 308489
个体 14	2. 942121	3. 402509	3. 304363	3. 67021	3. 249389	3. 318391
个体 15	3. 087126	3. 401329	3. 55029	3. 632719	3. 365708	3. 233357

从表 5.1 中可以看出，随着数据长度 L 增大，对同一个体，无论是睡眠期还是清醒期多参量生物电信号耦合程度是呈缓慢增长状态^[55]。不同个体多参量生物电信号耦合程度间的差异较大，这是由于不同个体处于不同的年龄段、体重等因素引起的。

将不同个体的清醒期及睡眠期生物电信号耦合程度进行平均、求最小值、最大值处理^[55]，结果如表 5.2 所示，并据此对睡眠期和清醒期的生物电信号耦合程度与数据长度 L 的关系进行对比分析。

表 5.2 睡眠期和清醒期平均耦合程度与数据长度 L 的关系对比

数据长度	L=500	L=1000	L=1500	L=2000	L=2500	L=3000
睡眠平均值	1.512794	1.560287	1.518343	1.56559	1.445454	1.508986
清醒平均值	1.837676	2.044646	2.195342	2.251975	2.220133	2.276874
睡眠最小值	0.607273	0.606597	0.894518	0.921369	0.820673	0.954379
清醒最小值	0.852019	0.773337	0.952402	0.865435	1.031482	1.257231
睡眠最大值	2.99702	2.639486	2.620442	2.420075	2.220133	2.47706
清醒最大值	3.087126	3.402509	3.55029	3.67021	3.613892	3.528177

分析表 5.2 数据能够得到以下结论^[55]:

- 1) 在数据程度 L 相同的情况下, 清醒生物电信号耦合程度大于睡眠生物电信号的。
- 2) 随着数据程度 L 的增长, 睡眠耦合程度平均值和清醒平均值基本保持不变。
- 3) 无论是清醒期还是睡眠期的生物电耦合程度个体差异较大。

睡眠期和清醒期平均耦合程度与数据长度 L 的关系更直观的对比如图 5.6 所示:

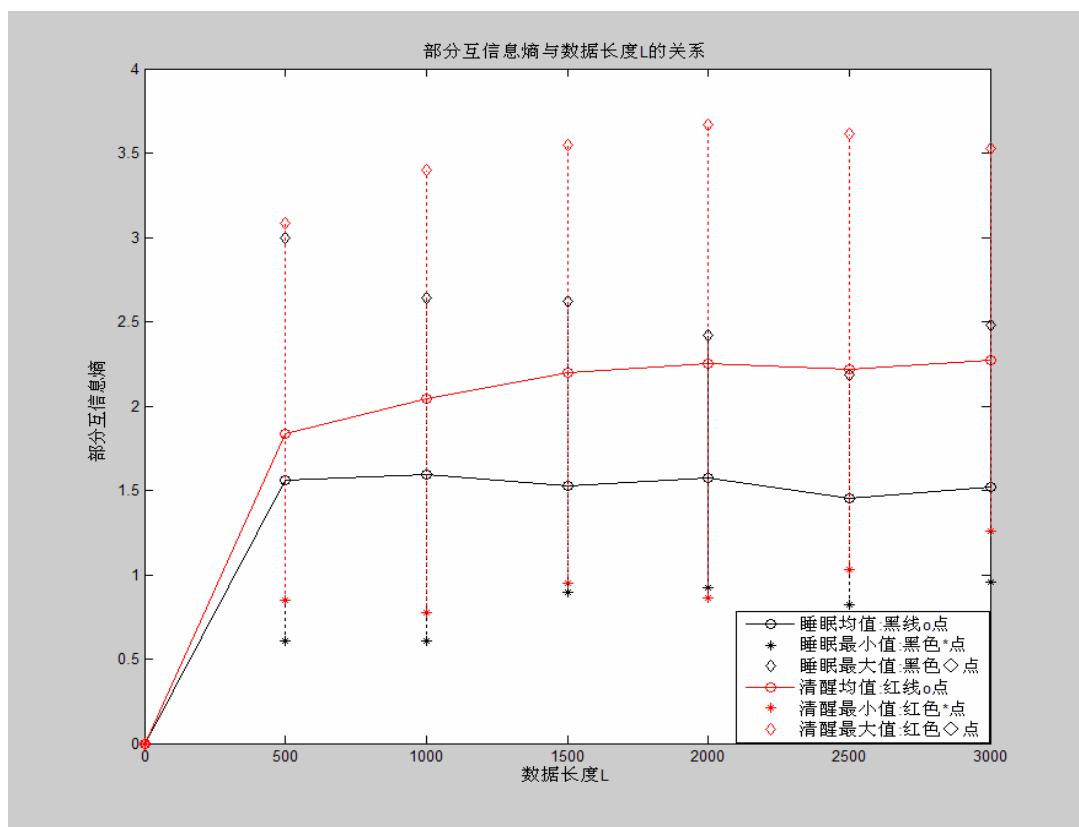


图 5.6 部分互信息熵与数据长度 L 的关系

分析图 5.6 睡眠期与清醒期的部分互信息曲线可知^[55]:

- 1) 在数据长度相等的情况下, 清醒期的生物电信号耦合程度要高于睡眠期。
- 2) 当数据长度由 0 增至 2000 过程中, 睡眠期与清醒期的耦合程度分别呈递增趋势^[1]。数据长度大于 2000 时, 耦合程度趋于平稳, 基本不变。

从数据的精准性考虑，数据长度越大，其统计概率分布越接近实际分布，相应的准确性越高^[55]。但从计算量以及计算速度来讲，数据长度越小，那么算法速度也越快，综合我们的实验结果，取数据长度 $L=2000$ 时，即能兼顾处理速度，又可以保证实验精度。

B. 研究多参量生物电信号耦合程度与编码长度 N 的关系

对受训者样本 15 组 Sleep 和 15 组 Weak 中的每组生物电信号序列分别取编码长度 $N=1,2,3,4,5,6$ ^[55]，并对每个时间序列进行符号化及编码处理，分别计算每个个体的部分互信息，并分别对睡眠期部分互信息和清醒期部分互信息进行平均，得到睡眠期及清醒期生物电信号耦合程度与编码长度 N 的关系如表 5.3 所示：

表 5.3 (a) 睡眠期生物电信号耦合程度与编码长度 N 的关系

编码长度	N=1	N=2	N=3	N=4	N=5	N=6
个体 1	0.00819	0.225973	0.770774	1.268618	1.47015	1.383171
个体 2	0.002158	0.114393	0.426473	0.826031	1.123609	1.230943
个体 3	0.0095	0.15939	0.574685	0.99869	1.243463	1.323501
个体 4	0.02291	0.183471	0.547126	0.92695	1.151599	1.230623
个体 5	0.011449	0.204937	0.780143	1.364247	1.651241	1.657733
个体 6	0.021699	0.126661	0.348737	0.521357	0.609059	0.656793
个体 7	0.009713	0.146827	0.555268	1.140849	1.769953	2.410442
个体 8	0.01077	0.145726	0.58612	1.092355	1.543939	1.91641
个体 9	0.004673	0.221189	0.812776	1.445948	2.004553	2.455185
个体 10	0.009996	0.084711	0.344209	0.713821	1.140422	1.599182
个体 11	0.004724	0.11387	0.42791	0.845763	1.293854	1.742791
个体 12	0.010146	0.150912	0.535103	0.96239	1.286376	1.475713
个体 13	0.002158	0.084711	0.344209	0.521357	0.609059	0.656793
个体 14	0.02291	0.225973	0.812776	1.445948	2.004553	2.455185
个体 15	0.01077	0.156339	0.555268	0.99869	1.293854	1.599182

表 5.3 (b) 清醒期生物电信号耦合程度与编码长度 N 的关系

编码长度	N=1	N=2	N=3	N=4	N=5	N=6
个体 1	0.020233	0.214107	0.741606	1.331154	1.746619	1.923064
个体 2	0.017981	0.190667	0.707733	1.367836	1.800912	1.913066
个体 3	0.015463	0.158677	0.499445	0.862724	1.112983	1.206347
个体 4	0.010317	0.125421	0.47082	0.902703	1.260396	1.447252
个体 5	0.010407	0.142908	0.555877	1.038805	1.441352	1.737011
个体 6	0.009855	0.154047	0.63162	1.213956	1.673445	1.868512
个体 7	0.008403	0.167871	0.733606	1.37223	1.88969	2.077471
个体 8	0.012209	0.16903	0.589933	1.030435	1.324281	1.430408
个体 9	0.004096	0.108548	0.525528	1.01099	1.514462	1.88791
个体 10	0.002853	0.209536	1.146399	2.298071	3.026081	3.23672
个体 11	0.004657	0.214212	0.987125	2.022938	2.738885	3.002279
个体 12	0.005145	0.169239	1.118924	2.372102	3.259076	3.539027
个体 13	0.00454	0.173331	0.967734	1.838696	2.408869	2.587599
个体 14	0.004907	0.176549	0.820231	1.717917	2.683983	3.558753
个体 15	0.008485	0.168468	0.764589	1.630498	2.564701	3.442331

从表 5.3 中可以看出，随着编码长度 N 从 2 增大到 5，对同一个体，无论是睡眠期还是清醒期多参量生物电信号耦合程度是呈线性增长状态，之后基本趋于平缓。不同个体多参量生物电信号耦合程度间的差异较大，这是由于不同个体处于不同的年龄段、体重等因素引起的^[55]。

将不同个体的清醒期及睡眠期生物电信号耦合程度进行平均、求最小值、最大值处理，结果如表 5.4 所示，并据此对睡眠期和清醒期的生物电信号耦合程度与编码长度 N 的关系进行对比分析^[55]。

表 5.4 睡眠期和清醒期的生物电信号所有个体平均耦合程度与编码长度 N 的关系

编码长度	N=1	N=2	N=3	N=4	N=5	N=6
睡眠平均值	0.009303	0.156339	0.561879	1.005309	1.350131	1.585319
清醒平均值	0.010785	0.169507	0.750745	1.467404	2.029716	2.32385
睡眠最小值	0.002158	0.084711	0.344209	0.521357	0.609059	0.656793
清醒最小值	0.002853	0.108548	0.47082	0.862724	1.112983	1.206347
睡眠最大值	0.02291	0.225973	0.812776	1.445948	2.004553	2.455185
清醒最大值	0.214212	0.214212	0.214212	0.214212	0.214212	0.214212

分析表 5.4 数据能够得到以下结论^[55]:

- 1) 在编码长度 N 相同的情况下, 清醒生物电信号耦合程度大于睡眠生物电信号的。
- 2) 随着编码长度 N 的增大, 睡眠耦合程度平均值和清醒平均值先是呈线性增长状态, 从 N=5 开始基本保持不变。
- 3) 无论是清醒期还是睡眠期的生物电耦合程度个体差异较大。

睡眠期和清醒期平均耦合程度与编码长度 N 的关系更直观的对比如图 5.7 所示:

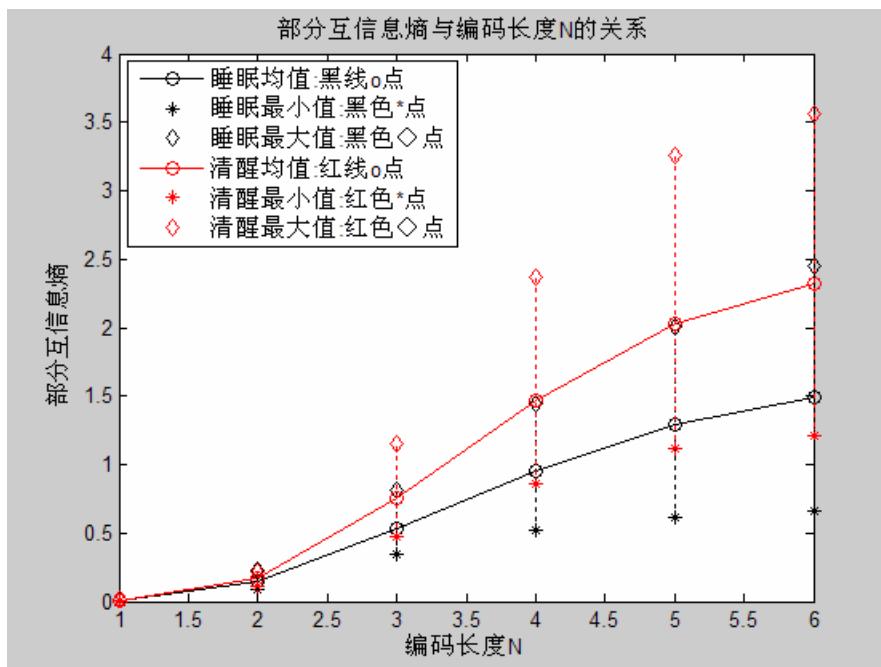


图 5.7 部分互信息熵与编码长度 N 的关系

分析图 5.7 部分互信息与编码长度 N 的关系可知^[55]:

- 1) 在编码长度 N 从 1 到 6 的递增过程中, 生物电信号的部分互信息都成增大趋势。
- 2) 清醒期的耦合程度明显高于睡眠期的, 且在 N=6 时清醒期与睡眠期耦合程度差异最显著^[1]。

考虑到实验效果的明显性, n 应该取大些; 考虑到算法的复杂度会影响实验处理速度以及在临床应用上的实时性, n 应该取小一些。综合考虑上述因素, n 取 6^[55]。

C.研究在数据长度 L=2000,编码长度 N=8 条件下清醒期睡眠期脑电信号耦合程度差异

对受训者样本 15 组 Sleep 和 15 组 Weak 中的每组生物电信号序列分别取编码长度 N=6 及数据长度 L=2000, 计算每个个体的部分互信息, 得到生物电耦合程度与睡眠期清醒期的关系如图 5.8 所示:

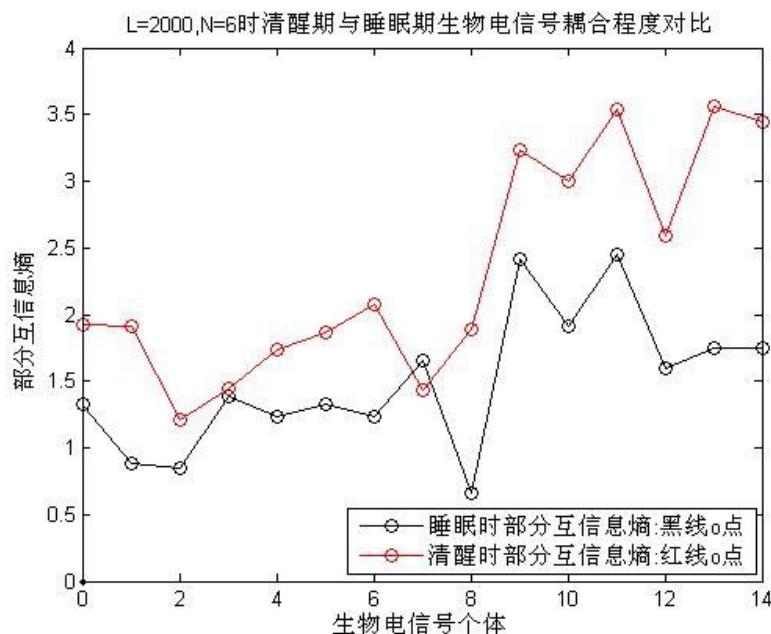


图 5.8 睡眠期与清醒期的生物电信号耦合程度对比

分析图 5.8 期与清醒期的生物电信号耦合程度对比^[55]可知:

- 1) 清醒期生物电信号耦合程度明显高于睡眠期的, 生物电信号的部分互信息熵可以作为衡量一个过程是否处于积极有序状态的参数, 耦合程度越高, 说明该物理过程越有序。
- 2) 上图中 0~14 号生物电信号个体间的生物电信号耦合程度有很大差异, 不同个体处于不同的年龄段和具有不同的体重引起的。

D.统计分析与假设检验

为进一步验证本文计算结果的准确性及算法的有效性^[55], 使用 spss 统计分析软件对计算结果进行了假设检验。主要方法如下:

对清醒期与睡眠期生物电信号耦合程度的差异显著性进行假设检验, 分别将睡眠期和清醒期的部分互信息值记作样本 S 和 W, 使用 spss 对两组部分互信息数据 S、W 进行独立样本 T 检验^[55], 结果如表 5.5 所示:

表 5.5 清醒期与睡眠期生物电信号耦合程度的差异显著性假设检验

组统计量				
样本	N	均值	标准差	均值的标准误
清醒	15	2.3239	0.8265	0.2134
睡眠	15	1.4935	0.5214	0.13464

独立样本 T 检验						
样本		方差的 Levene 检验		均值方程的 t 检验		
		F	Sig.	t	df	Sig. (双侧)
清醒 睡眠	假设方差相等	6.866	0.014	3.291	28	0.003
	假设方差不相等	6.866	0.014	3.291	28	0.003

样本		均值方程的 t 检验			
		均值差值	标准误差值	差分的 95% 置信区间	
				下限	上限
清醒 睡眠	假设方差相等	0.8303322637	0.2523197945	0.3091272950	1.3515372323
	假设方差不相等	0.8303322637	0.2523197945	0.3134785945	1.3471859328

本文关注的是准确的对清醒期与睡眠期生物电信号进行判断和辨别^[1], 根据表 5.5 分析清醒期与睡眠期生物电信号耦合程度的差异显著性, 假设清醒期与睡眠期的均值相等:

①由表 5-5 可知, 在假设方差相等的情况下, 自由度(df)为 28, 查找 t 值表可知理论 t 值 $t(df)0.05=t(28)=2.048$, 而样本 t 值为 3.291, 大于 $t(28)$, 且 $Sig(\text{双侧})=0.003<0.05$, 假设不成立. 所以清醒期与睡眠期生物电信号耦合程度差异显著。

②由表 5-5 可知, 在假设方差不相等的情况下, 自由度(df)为 22, 查找 t 值表可知理论 t 值 $t(df)0.05=t(23)=2.048$, 而样本 t 值为 2.484, 大于 $t(22)$, 且 $Sig(\text{双侧})=0.003<0.05$, 假设不成立. 所以清醒期与睡眠期生物电信号耦合程度差异显著。证明该算法可以有效的对清醒期与睡眠期生物电信号进行区分。

以上实验结果证明, 使用符号化的部分互信息算法可以很好的对多参数生物电信号的耦合程度进行分析, 以生物电信号各参量之间的耦合程度作为参数, 能够有效的辨别该生理机制是否是活跃的, 这对于身体内部器官健康状况的诊断具有重要意义^[55]。

5.4 小结

本章首先介绍信息熵研究现状，以及互信息及部分互信息基本原理。接着介绍了基于符号化部分互信息算法的时间序列耦合分析算法基本理论，经验证该算法能够很好的分析离散随机变量间的耦合信息^[55]。在此基础之上，使用符号化的部分互信息来计算生物电时间序列间的耦合程度，实现了利用统计学来量化研究生命现象。基于该算法分别对睡眠期和清醒期的 EEG、ECG、EMG 多参量生物电信号进行分析，分析结果表明，清醒期的生物电信号的耦合程度显著高于睡眠期的。并进行了假设验证，证明睡眠期和清醒期的生物电耦合信息具有显著差异，表明耦合程度可以作为衡量一个物理过程是否处于积极状态以及睡眠分期的参数^[55]。

第六章 总结与展望

生物体内的生物电现象是一种正常的生理活动，一定的生理活动对应于一定的生物电反应，即生物电在统计意义上存在一定的规律性。因此，从统计学角度来分析生物电，可以为我们提供一定的生理活动的信息，据此可以判断一个生理机制是否处于健康有序的状态。目前生物电的研究已经广泛应用于临床医学，如脑电图、心电图、肌电图等等，医务工作人员可以根据这些信息来诊断病情。在可预见的未来，生物电在仿生学、信息控制、物联网等领域也会产生巨大的作用。鉴于此，本文的工作主要有以下三个方面：

一、青老年脑电信号时间不可逆的分析及对比

针对传统算法的不足并且根据脑电信号的特点，提出了改进的符号化相对熵算法，该算法可以对生物电信号的不可逆性进行分析。该算法是一种新的使用过程的前向概率和后向概率计算符号相对熵^[1]，并利用符号相对熵来估计熵产的方法。该方法是基于熵增和过程不可逆特性关系的，本文同时证明脑电信号具有时间不可逆特性，而且该不可逆特性可以提供脑电信号的熵增信息。最后应用该方法对青老年脑电信号进行数值计算及对比，指出了青老年人的脑电信号的不可逆性存在差异，结果是老年人的平均能量损耗显著高于年轻人。并给出了青老年符号相对熵值的变换范围，能够为脑电活动生理过程算法处于积极有序状态的判断提供依据，证明符号相对熵可以作为一个物理过程不可逆程度的度量参数^[1]，这对脑电信号是否处于积极或平衡状态的诊断治疗具有积极的作用，未来可应用于生命状态分析、临床医学、生物电传感器等领域。

二、癫痫及正常脑电信号不可逆性分析对比

应用基于符号化相对熵算法法对癫痫异常脑电信号及正常脑电信号进行数值计算及对比，结果是癫痫患者的平均能量损耗显著高于正常人的，证明符号相对熵可以作为衡量脑电信号是否异常的参数。

三、多参量生物电信号的耦合分析

针对传统互信息分析方法的缺陷以及算法的复杂性，提出了新的生物电信号耦合程度分析方法—符号化部分互信息熵。这种方法是基于部分互信息的。主要方法是首先对原始序列进行符号化，对符号化的序列提取部分互信息。针对 matlab 处理循环效率低下的不足，算法的概率统计部分采用了 C 和 matlab 编程结合的方式。本文同时证明，多参量的生物电信号各参量间具有耦合关系，使用符号化的部分互信息能够很好的对生物电信号时间序列进行分析，从而获得其耦合程度。应用该算法对生物电信号计算并进行假设检验，结果表明清醒期的生

物电信号耦合程度显著高于睡眠期的，证明符号化部分互信息可以用来分析时间序列间的耦合信息，而且生物电信号的耦合程度可以作为度量一个物理过程是否处于活跃状态的参数，未来该算法未来可用于人体工作和休息效率等状态的改善等领域。

本文提出的两个算法对于实际问题的解决具有重要意义，但是符号化相对熵算法的耗时还存在改进的空间。因此在下一步工作中，一是要缩短算法的时间，提高算法的收敛速度，更加符合临床使用的要求。二是要从实际运用的角度出发，例如在临床过程中信号数目多于信号导联数的病态问题时，提高算法对感兴趣信号的提取能力。另外，本文中提出的算法也可以应用于其他更广泛的领域，比如符号化相对熵算法可以用于分析任何时间过程或变化系统的不可逆程度。而符号化部分互信息系统可以应用于多个耦合系统间相关性的分析及研究。

参考文献

- [1] 张梅,王俊.基于改进的符号相对熵的脑电信号时间不可逆性研究[J].物理学报,2013,62(3):038701-038707.
- [2] 宁新宝,卞春华,王俊等.心脏电活动过程的非线性分析[J].科学通报,2006,51(7):764-771.
- [3] RAJENDRA A U,MOLINARIB B,VINITHS S S.Automated diagnosis of epileptic EEG using entropies [M].Biomedical Signal Processing and Control,2012:401-408.
- [4] 郭继鸿,张萍.动态心电图学[M].北京:人民卫生出版社,2003:615-683.
- [5] 王春光.基于稀疏分解的心电信号特征波检测及心电数据压缩[D].长沙:国防科技大学,2009:14-15.
- [6] KURTHS J,VOSS A ,SAPARIN P,et al.Quantitative analysis of heart rate variability[J].Chaos,1995,5(1): 88-94.
- [7] CHEN Z,IVANOV P C,HU K,et al.Effect of nonstationarities on detrended fluctuation analysis[J]. Physical Review E,2002,65:041107-041110.
- [8] 杨琦.时间压力对脑电信号特征的影响[D].上海:上海交通大学,2010.
- [9] HAO B L.Symbolic dynamics and characterization of complexity[J].Physica D,1991,51(2):161-176.
- [10] 张辉,杨明静.非线性动力学在心脏活动研究中的应用[J].生物物理学报,1997,13(2): 340 -346.
- [11] LI J,NING X B,WU W,et al.Detecting dynamical complexity changes in time series using the base-scale entropy[J].Chinese Physics,2005,14(12):2428-2432.
- [12] COSTA M,GOLDBERGER A L,PENG C K.Multiscale entropy analysis of complex physiologic time series[J].Physical Review Letters,2002,89(6):0681021-0681024.
- [13] ROSSO O A,LARRONDO H A,MARTIN M T.Distinguishing Noise from Chaos[J].Physical Review Letters,2007,98 (2):99-103.
- [14] 蔡琼.基于小波变换的EEG分析[D].长沙:长沙理工大学.2007.
- [15] 宁新宝编著.生物医学电子学[M].长沙:湖南科学技术出版社,1988:17-19.
- [16] 罗崇.基于nRF24E1的大鼠脑电遥测系统的研究[D].成都:电子科技大学,2009.
- [17] 王振洲,宁新宝,张宇,等.同步十二导联心电图信号的关联维分布[J].科学通报,2000,45(8):873-877.
- [18] MALLAT S,WEN H L.Singularity detection and processing with wavelet [J].Institute of electrical and electronics engineers transaction on information technology,1992,38 (2):617-643.
- [19] LI J, NING X B.Dynamical complexity detection in short-term physiological series using base-scale entropy[J].Physics Review E,2006, 45(8):73-76.
- [20] 尚文霞.重剑基本动作和专项力量训练肌肉用力特征肌电分析[D].北京:北京体育大学,2008,04.
- [21] 沈凤麟,李山河,梁岩.心电地图仪中工频干扰的一种滤除方法[J].数据采集与处理,1994, 9(3):183-189.
- [22] 孙文铎.十二导联心电信号采集及分析系统的研究[D].吉林:吉林大学,2006.
- [23] 张翔,周严.基于自适应滤波的心电图中呼吸干扰的抑制[J].生物医学工程学杂志,2011,28(1):58-62.
- [24] ANDRIEUX D,GASPARD P.Entropy Production and Time Asymmetry in Nonequilibrium Fluctuations[J]. Physical Reciew Letters,2007,45(8):98-101.
- [25] LARRONDO H A,GONZALEZ C M,MARTIN M T,et al.Intensive statistical complexity measure of pseudorandom number generators[J].Physica A,2005,356(9):133-138.
- [26] 宋爱玲,黄晓林,司峻峰,等.符号动力学在心率变异性分析中的参数选择 [J].物理学报,2011,60(2):126-133.
- [27] 李锦,宁新宝,马千里.用联合熵分析短时心率变异信号的非线性动力学复杂性[J].生物医学工程学杂志,2007,24(2):285-289.

- [28] GOLDBERG A L,AMARAL L,GLASS L,et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220
- [29] 林欣.十二导联心电数据库生成系统的设计与实现[D]. 青岛:中国海洋大学,2007.
- [30] EDGAR R,JUAN M R.Estimating Dissipation from Single Stationary Trajectories[J].Physical Review Letters,2010,45(8):105-108.
- [31] 杨水祥,胡大一.心血管热点荟萃 [M].北京:人民卫生出版社,2011,22(7):1-2.
- [32] GRASSBERGER P,PROCCACCIA I.Estimation of the kolmogorov entropy from a chaotic signal[J].Physical Review,1983,28(4):2591-2593.
- [33] COSTA M,GOLDBERGER A L,PENG C K.Broken Asymmetry of the Human Heartbeat: Loss of Time Irreversibility in Aging and Disease[J].Physical Review Letter,2005,95(19):1981021-1981024.
- [34] PORTA A,CASALI K R,CASALI A G,et al.Temporal asymmetries of short-term heart period variability are linked to autonomic regulation[J].American Journal of Physiology,2008,295(1):1-29.
- [35] CAMILLO C,ENRICO R,Time reversal. Symbolic series and irreversibility of human heartbeat[J].Chaos, Solitons and Fractals,2007,32(5):1649-1654.
- [36] COSTA M,PENG C K,GOLDBERGER A L.Multiscale Analysis of Heart Rate Dynamics: Entropy and Time Irreversibility Measures[J].Cardiovascular Engineering ,2008,8(1):88-93.
- [37] DAW C S,FINNEY C E A,TRACY E R.A review of symbolic analysis of experimental data[J].Review of Scientific Instruments,2003,74(3): 915-933.
- [38] STEUER R,MOLGEDEY L,EBELING W,et al.Entropy and optimal partition for data analysis[J].The European Physical Journal B,2001,19(5): 265-269.
- [39] 李锦,宁新宝. 短时心率变异性信号的基本尺度熵分析[J].科学通报,2005,50(14):24-28.
- [40] WESSEL N,ZIEHMANN C,KURTHS J,et al.Short-term forecasting of life-threatening cardiac arrhythmias based on symbolic dynamics and finite-time growth rates[J].Physical Review E.2000,61(1):733-739.
- [41] 李金坤.颞叶癫痫大鼠认知功能改变与脑电的相关性研究[D].成都:电子科技大学,2011.
- [42] 卞春华,马千里,司俊峰等. 短时心率变异符号序列的熵分析方法[J].科学通报,2009,54(3):340-344.
- [43] 黄晓林,崔胜忠,宁新宝等. 心率变异性基本尺度熵的多尺度化研究[J].物理学报,2009,58(12):8160-8165.
- [44] THOMAS M C,JOY A T.信息论基础[M].北京,机械工业出版社,2007:13-14.
- [45] PORPORTAO A,RIGBY J R,DALY E.Irreversibility and Fluctuation Theorem in Stationary Time Series[J].Physical Review Letters,2007,98(3):094101-094104.
- [46] 马斌荣,陈卉.医学科研中的统计方法[M].北京,科学出版社,2005, 44-45.
- [47] 祝怀珍.加强胎儿心电监护提高新生儿成活率[J]. 实用心电学杂志,2006,15(5):335-338.
- [48] 沈麟,王俊.基于符号相对熵的心电信号时间不可逆性分析[J].物理学报,2011,60(11):118702-118706.
- [49] 李文举,梅丽,信润海,等.基于 KL 散度和 BP 神经网络的人类基因启动子识别[J].辽宁师范大学学报,2010,33(1):42-45.
- [50] ZHANG M,WANG J,CUI C,et al.Epilepsy abnormal electroencephalogram analysis based on symbolic relative entropy[C],2013 International Conference on Mechatronics and Industrial Informatics.Accepted).
- [51] STEFAN F,BERND P.Partial Mutual Information for Coupling Analysis of Multivariate Time Series[J], . Physical Review Letters,2007,99(4):204101-204104.
- [52] RECHTSCHAFFEN A, KALES A.A Manual of Standardized Terminology,Techniques and Scoring System for Sleep Stages of Human Subjects[J].National Institute of Health Publication,1969.
- [53] STAM C J.Nonlinear dynamical analysis of EEG and MEG review of an emerging field [J].Clinical Neuro-physiology.2005.116 (10):2266-2301.
- [54] RICHMAN J S,MOON R J.Physiological time-series analysis using approximate entropy and sample entropy[J],Am J Physiol Heart Circ Physiol,2000,278(6): 2039-2049.
- [55] 张梅,崔超,马千里,等,基于符号化部分互信息熵的多参数生物电信号的耦合分析[J].物理学报,2013,62(6):068704-068709.

- [56] ZHON W X.Multifractal detrended cross-correlation analysis for two nonstationary signals[J].Physical Review E,2008(5),77: 1-4.
- [57] CHIL L P,YANG C B,MA K,et al.Detrended fluctuation analysis on correlations of complex networks under attack and repair strategy[J].Physics,2006, 45(5): 765-768.
- [58] FERRI R,ELIA M,MUSUMECI S A,et al.The time course of frequency bands (15-45 Hz) in all-night spectral analysis of sleep EEG [J]. Clinical Neurophysiology 2000, 111(7):1258- 1265.
- [59] MUKAI J,UCHIDA S.Spectral analysis of all-night human sleep EEG in narcoleptic patients and normal subjects [J].European Sleep Research Society J. Sleep Res, 2003, 12(7):63-71.
- [60] MERICAL H,RONALD D.Fortune spectral power time-courses of human sleep EEG reveal a striking discontinuity at 18 Hz marking the division between NREM-specific and Wake/REM-specific fast frequency activity[J],Cerebral Cortex Advance Access ,2005,15(7):877-884.
- [61] ACHARYA U R,FAUST O, KANNATHAL N.Non-linear analysis of EEG signals at various sleep stages [J]. Compute Methods Programs Biomed, 2005, 80(1):37-45.
- [62] NAX,PENG S,ANTI KAMAE.Modeling traffic flow correlation using DFA and DCCA [J]. Springer Science Business Media B.V. 2010, 80(1):11071-11075.
- [63] MALLAT S.Characterization of Signals From Multiscale Edges [J]. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 1992, 14(7):710-732.
- [64] MUZY J F,BACRY E,AMEODO A.Multifractal formalism for fractal signals:The structure-function approach versus the wavelet-transform modulus-maxima method [J].Physical Review E,1993, 47(2):875-884.
- [65] KESTENER P,AMEODO A.Generalizing the wavelet-based multifractal formalism to random vector fields: application to three-dimensional turbulence velocity and vorticity data [J].Physical Review Letters, 2004, 93(4):45011-45014.
- [66] PODOBNIK B,STANLEY H E.Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series [J].Physical Review Letters,2008, 100(7):0841021-0841024.
- [67] ZEBENDE G F,MACHADO A.Cross-correlation between time series of vehicles and passengers [J].Physica A 2009,388(7):4863-4866.
- [68] ERINALDO L S,TATIJANA S,LUCIAN B,et al.Correlations and cross-correlations in the Brazilian agrarian commodities and stocks [J].Physica A, 2010, 389(4):2739-2743.
- [69] HAJIAN S, SADEGH M.Multifractal detrended cross-correlation analysis of sunspot numbers and river flow fluctuations [J].Physica A, 2010, 389(21):4942-4957.
- [70] SHADKHOO S,JAFARI G R.Multifractal detrended cross-correlation analysis of temporal and spatial seismic data[J],The European Physical Journalal B,2009,72(4): 679–683.
- [71] PLAMEN C L,QIAN D Y,RONNY P B.Maternal–fetal heartbeat phase synchronization[J],Proceeding of the National Academy of Sciences ,2009,33(2):13641-13642.
- [72] WANG J,MA Q L.Multiscale entropy based study of the pathological time series[J].Chines Physics B,2008, 72(7):4424-4427.
- [73] 许小可.海杂波的非线性分析与建模[D].大连:大连海事大学,2007.
- [74] PEDRO W L,ANA P M.Non-logarithmic Jensen–Shannon divergence[J]. Physica A,2003,329:81-90.
- [75] BASHAN A,BARTSCH R P,KANTELHARDT J W. Network physiology reveals relations between network topology and physiological function.Nature Communications,2012,10(4):1038-1040.
- [76] PARK K L,LEE K J.Application of a wavelet adaptive filter to minimise distortion of the ST-segment[J]. Medical & Biological Engineering & Computing,1998,36(5):97-100.
- [77] 赵大庆,王俊.小波多重分形在脑电信号分析中的应用[J].中国生物医学工程学报,2010,29(5):793-796.
- [78] SHANNON C E,WEAVER W. The Mathematical Theory of Information[M].University of Illinois Press, Urbana, IL,1949, 36(7):136-142.

-
- [79] WANG J,YU Z F.Symbolic transfer entropy-based premature signal analysis[J].Chinese Physical Society and IOP Publishing Ltd,2012,36(5):018702-018705.
 - [80] 曹雪虹,张宗橙编著.信息论与编码[M],北京:清华大学出版社,2004.3.23-30.
 - [81] BANDT C,POMPE B.Permutation entropy: a natural complexity measure for time series[J].Physical Review Letters, 2002, 88(9): 1741- 1743.
 - [82] THOMAS S.Measuring Information Transfer[J].Physical Review Letters,2000, 01870(7):461-463.
 - [83] STANIEK M,LEHNERTZ K,Symbolic transfer entropy[J].Physical Review Letters,2008,100(3): 1581011-1581014.
 - [84] 曹寅文,宋慎义,肖井华.运动后人体心肺节律同步关系及信号的耦合方向 [J].物理学报.2010,7(59),5164-5167.
 - [85] BAHRAMINASAB A,GHASEMI F,STEFANOVSKA A,et al.Direction of Coupling from Phases of Interacting Oscillators:A Permutation Information Approach[J].Physical Review Letters,2008,100(7): 0841011-080414.
 - [86] CAMPILLO M, PAULA A.Long-range correlation in the diffuse seismic coda[J].Science,2003,299(6): 547-549.
 - [87] PLEROU V,GOPIKRISHNAN P,ROSENOW B,et al.Universal and Non-universal Properties of Cross correlations in Financial Time Series[J].Physical Review Letters,1999,83(7):1471-1474.
 - [88] 胡细宝,孙洪祥.概率论与随机过程[M].北京:北京邮电大学出版社,2001: 213-223.

附录 1 攻读硕士学位期间撰写的论文

- (1) 张梅,王俊.基于改进的符号相对熵的脑电信号时间不可逆性研究[J].物理学报,2013,62(3):038701.
- (2) 张梅,崔超,马千里,等.基于符号化部分互信息熵的多参数生物电信号的耦合分析[J].物理学报,2013,62(6):068704.
- (3) Mei Zhang, Junchang Zhao, Zhengzhong Zheng, Jiafei Dai, Jun Wang. Epilepsy abnormal electroencephalogram analysis based on symbolic relative entropy[], 2013 International Conference on Mechatronics and Industrial Informatics. (Accepted)

附录 2 攻读硕士期间参加项目

- (1) 国家自然科学基金面上项目，基于去趋势互相关和符号转移熵的心脑电耦合研究(61271082);
- (2) 国家自然科学基金青年基金，睡眠节律瞬态特性研究与缺省网络构建，(61201029);
- (3) 江苏省科技厅，基于缺省网络的睡眠节律瞬态特性研究，(BK2011759)。