

面向康复工程的脑电信号分析和判别模型

● 背景和意义

大脑是人体中高级神经活动的中枢，拥有着数以亿计的神经元，并通过相互连接来传递和处理人体信息。脑电信号按其产生的方式可分为诱发脑电信号和自发脑电信号。诱发脑电信号是通过某种外界刺激使大脑产生电位变化从而形成的脑电活动；自发脑电信号是指在没有外界特殊刺激下，大脑自发产生的脑电活动。

（1）诱发脑电信号（P300 脑-机接口）

在日常生活中，人的大脑控制着感知、思维、运动及语言等功能，且以外围神经为媒介向身体各部分发出指令。因此，当外围神经受损或肌肉受损时，大脑发出指令的传输通路便会受阻，人体将无法正常工作完成大脑指令的输出，也就失去了与外界交流和控制的能力。研究发现，在外围神经失去作用的情况下，人的大脑依旧可以正常运行，而且其发出指令的部分信息可以通过一些路径表征出来。脑-机接口技术旨在不依赖正常的由外围神经或肌肉组织组成的输出通路的通讯系统，实现大脑与外部辅助设备之间的交流沟通。

P300 事件相关电位是诱发脑电信号的一种，在小概率刺激发生后 300 毫秒范围左右出现的一个正向波峰（相对基线来说呈现向上趋势的波）。由于个体间的差异性，P300 的发生时间也有所不同，图 1 表示的是在刺激发生后 450 毫秒左右的 P300 波形。P300 电位作为一种内源性成分，它不受刺激物理特性影响，与知觉或认知心理活动有关，与注意、记忆、智能等加工过程密切相关。基于 P300 的脑-机接口优点是使用者无需通过复杂训练就可以获得较高的识别准确率，具有稳定的锁时性和高时间精度特性。

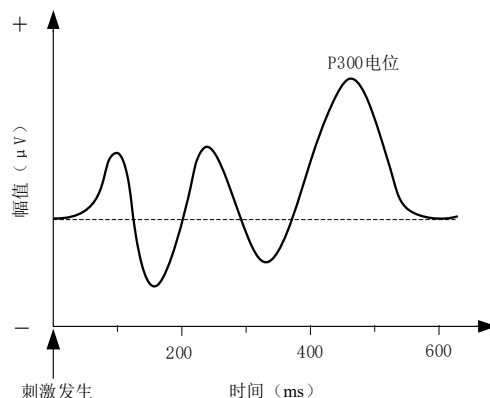


图 1 P300 波形示意图

(2) 自发脑电信号（睡眠脑电）

睡眠是身体休整积蓄能量的重要环节，睡眠质量对人的身心状态也有着重大影响。如何提高睡眠质量，减少睡眠相关疾病对健康的影响，日益受到广泛关注。睡眠过程中采集的脑电信号，属于自发型的脑电信号。自发型的睡眠脑电信号能够反映身体状态的自身变化，也是用来诊断和治疗相关疾病的重要依据。

睡眠过程是一个动态变化的复杂过程。在国际睡眠分期的判读标准 R&K 中，对睡眠过程中的不同状态给出了划分：除去清醒期以外，睡眠周期是由两种睡眠状态交替循环，分别是非快速眼动期和快速眼动期；在非快速眼动期中，根据睡眠状态由浅入深的逐步变化，又进一步分为睡眠 I 期，睡眠 II 期，睡眠 III 期和睡眠 IV 期；睡眠 III 期和睡眠 IV 期又可合并为深睡眠期。图 2 给出了不同睡眠分期对应的脑电信号时序列，自上而下依次为清醒期、睡眠 I 期、睡眠 II 期、深睡眠和快速眼动期。从图 2 中可以观察到，脑电信号在不同睡眠分期所呈现的特点有所不同。基于脑电信号进行自动分期，能够减轻专家医师的人工负担，也是评估睡眠质量、诊断和治疗睡眠相关疾病的重要辅助工具。

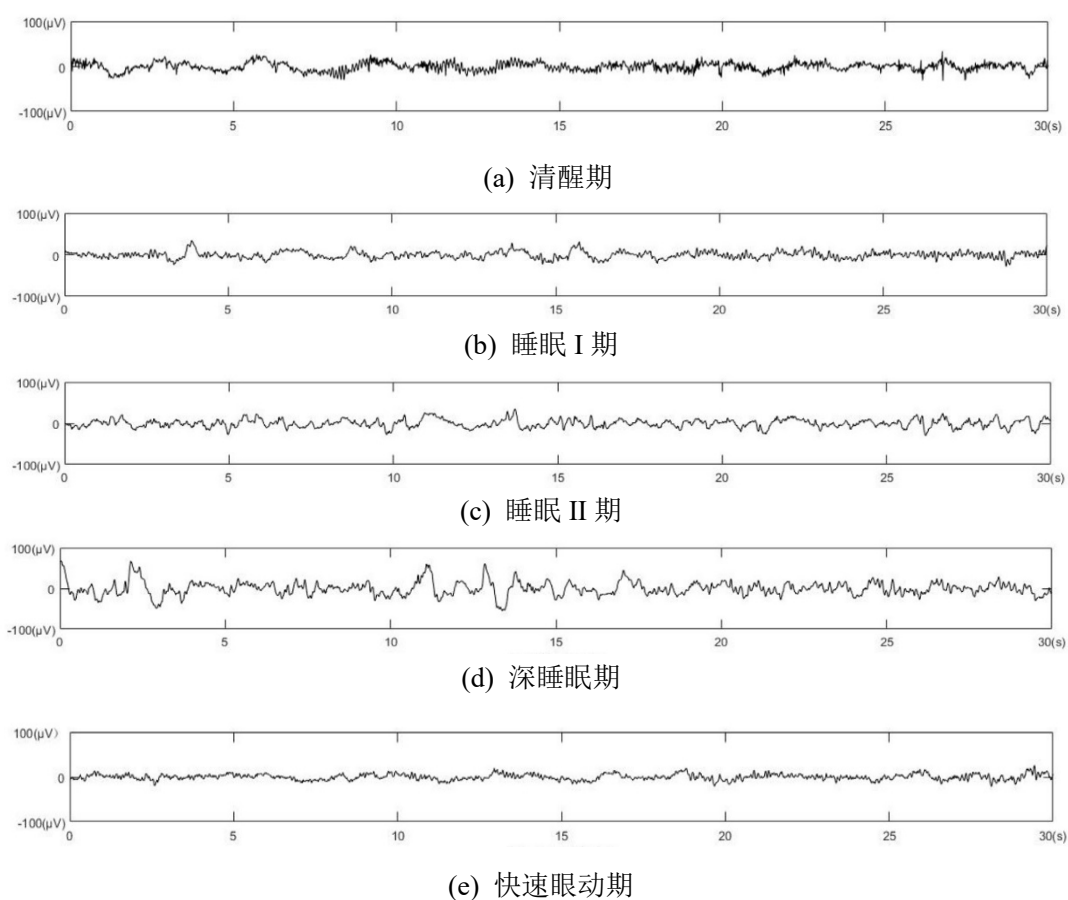


图 2 各睡眠分期的睡眠脑电信号时序列

● 课题任务

本赛题包含 2 个附件（数据文件），四个课题任务。具体说明如下，

附件 1：P300 脑机接口实验数据

提供了 5 个健康成年被试（S1-S5）的 P300 脑机接口实验数据，平均年龄为 20 岁。在实验的过程中，要求每一位被试（被测试者）集中注意力。P300 脑机接口实验的设计如下：每位被试能够观察到一个由 36 个字符组成的字符矩阵，如图 3 所示，字符矩阵以行或列为单位（共 6 行 6 列）。每轮实验的设计流程：首先，提示被试注视“目标字符”，例如在图 3 的字符矩阵上方，出现的灰色字符“A”；其次，进入字符矩阵的闪烁模式，每次以随机的顺序闪烁字符矩阵的一行或一列，闪烁时长为 80 毫秒，间隔为 80 毫秒；最后，当所有行和列均闪烁一次后，则结束一轮实验。在被试注视“目标字符”的过程中，当目标字符所在行或列闪烁时，脑电信号中会出现 P300 电位；而当其他行和列闪烁时，则不会出现 P300 电位。上述实验流程为 1 轮，共重复 5 轮。

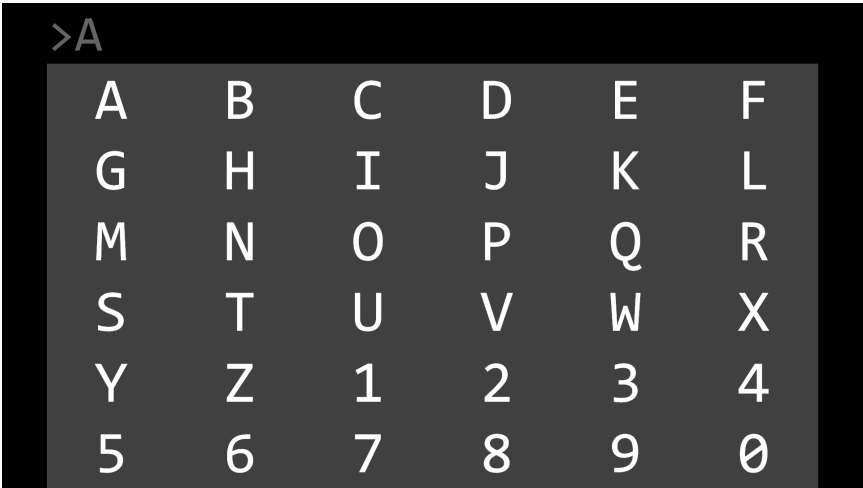


图 3 字符矩阵界面

每位被试的 P300 脑电数据包含有 4 个文件，具体说明如下，

train_data：训练用数据；

train_event：训练数据的事件标签；

test_data：测试用数据；

test_event：测试数据的事件标签。

训练用数据包括 12 个已知目标字符的数据（char01~char12），测试用数据包括 10 个待识别目标字符的数据（char13~char22）。每个字符矩阵闪烁实验中，脑

电数据表格包含有 20 列（每列表示 1 个记录通道，记录通道依次进行编号，表 1 为记录通道的标识符，图 5 对应了记录通道的位置），脑电数据表格的行表示样本点数据，采样频率为 250Hz。信号采集设备设置了参考电极和接地电极，即记录通道的信号为作用电极与参考电极之间的差值。

表 1 采集通道的标识符

图 5 脑电信号采集通道图

测试数据中的标签文件同样是以子表形式与实验数据相对应，子表的名称为“charXX”，XX 对应相应字符的序列号。在测试数据的事件标签文件中，第一行给出了待识别目标字符的标识符，统一表示为“666”，需要通过对脑电信号进行分析后，得到出现 P300 电位的行和列，并判断得到目标字符的识别结果。

A	101	B	102	C	103	D	104	E	105	F	106
----------	-----	----------	-----	----------	-----	----------	-----	----------	-----	----------	-----

G	107	H	108	I	109	J	110	K	111	L	112
M	113	N	114	O	115	P	116	Q	117	R	118
S	119	T	120	U	121	V	122	W	123	X	124
Y	125	Z	126	1	127	2	128	3	129	4	130
5	131	6	132	7	133	8	134	9	135	0	136



图 6 行/列的标识符

附件 2：睡眠脑电数据

提供 3000 个睡眠脑电特征样本及其标签，取自不同的健康成年人整夜睡眠过程。第一列为“已知标签”，用数字形式来表示不同的睡眠分期：清醒期（6），快速眼动期（5），睡眠 I 期（4），睡眠 II 期（3），深睡眠期（2）；第二至五列为从原始时序列中计算得到的特征参数，依次包括“Alpha”，“Beta”，“Theta”，“Delta”，分别对应了脑电信号在“8-13Hz”，“14-25Hz”，“4-7Hz”和“0.5-4Hz”频率范围内的能量占比，特征参数单位为百分比。

根据以上附件所给出的数据来源和实验数据，请研究以下问题：

问题一：在脑-机接口系统中既要考虑目标的分类准确率，同时又要保证一定的信息传输速率。请根据附件 1 所给数据，设计或采用一个方法，在尽可能使用较少轮次（要求轮次数小于等于 5）的测试数据的情况下，找出附件 1 中 5 个被试测试集中的 10 个待识别目标，并给出具体的分类识别过程，可与几种方法进行对比，来说明设计方法的合理性。

问题二：由于采集的原始脑电数据量较大，这样的信号势必包含较多的冗余信息。根据图 5 和表 1，在 20 个脑电信号采集通道中，无关或冗余的通道数据不仅会增加系统的复杂度，且影响分类识别的准确率和性能。请分析附件 1 所给数据，并设计一个通道选择算法，给出针对每个被试的、更有利于分类的通道名

称组合（要求通道组合的数量小于 20 大于等于 10，每个被试所选的通道可以不相同，具体的通道名称见图 5 和表 1）。基于通道选择的结果，进一步分析对于所有被试都较适用的一组最优通道名称组合，并给出具体分析过程。为了方便参赛者对最优通道组合进行选择，赛题给出了测试数据（char13-char17）的结果，它们的字符分别是：M、F、5、2、I。

问题三：在 P300 脑-机接口系统中，往往需要花费很长时间获取有标签样本来训练模型。为了减少训练时间，请根据附件 1 所给数据，选择适量的样本作为有标签样本，其余训练样本作为无标签样本，在问题二所得一组最优通道组合的基础上，设计一种学习的方法，并利用问题二的测试数据（char13-char17）检验方法的有效性，同时利用所设计的学习方法找出测试集中的其余待识别目标（char18-char22）。

问题四：根据附件 2 中所给出的特征样本，请设计一个睡眠分期预测模型，在尽可能少的训练样本的基础上，得到相对较高的预测准确率，给出训练数据和测试数据的选取方式和分配比例，说明具体的分类识别过程，并结合分类性能指标对预测的效果进行分析。

注意：在研究问题一时，不能将问题二提供的 5 个测试数据的结果作为已知条件，若使用该条件，则被判为作弊处理。