

听觉通道语音冲突大脑皮层电位的 听觉认知控制特征提取方法*

于 波^{1,2} 李海峰¹ 马 琳¹ 王勋达¹

(1 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

(2 哈尔滨理工大学软件学院 哈尔滨 150040)

2015 年 7 月 19 日收到

2016 年 4 月 1 日定稿

摘要 认知心理学发现, 视觉、听觉接收到信息有冲突时, 大脑皮层电位会发生扰动, 由此可探索认知冲突控制的“刺激-反应”机制。视觉认知冲突实验较多, 成果丰硕, 而相应的听觉实验很少, 并且得到不一样的结论。本研究利用冲突和非冲突的语音信号刺激, 分析研究脑电信号, 提出基于三阶段听觉认知控制的时域特征模型。研究人脑听觉通道在出现语音认知冲突时的认知控制的规律下的单次试验脑电数据特征提取方法。根据得到的认知规律, 单次试验脑电样本被分成 3 个部分。被分割的每个阶段使用时域上的平均幅值和 Lempel-Ziv 复杂度 (LZC) 进行计算, 从而联合 3 个阶段的特征作为听觉认知脑电样本的特征。结果表明: (1) 先发现的认知冲突相关的混合脑电成分“N1-P2&N2&Late-SW”分别体现了听觉认知控制的 3 个阶段; (2) 一个更完整的听觉认知控制过程应包括 3 个阶段的时域特征: 感知阶段: 110~140 ms, 识别阶段: 260~320 ms, 解决阶段: 500~700 ms; (3) 提出针对单次听觉认知控制脑电样本的特征提取方法, 联合使用平均幅度和 LZC 可以获得最好的识别率 (99.33%)。实验结果证明了提出的方法能够有效地检测听觉认知控制脑电数据, 进而提供人脑认知控制能力评价的声学方法。

PACS 数: 43.60, 43.66, 43.71, 43.72, 43.80

The method of auditory cognitive control feature extraction of cerebral cortex potential for auditory modality speech conflicting

YU Bo^{1,2} LI Haifeng¹ MA Lin¹ WANG Xunda¹

(1 School of Computer Science and Technology, Harbin Institute of Technology Harbin 150001)

(2 School of Software, Harbin University of Science and Technology Harbin 150040)

Received Jul. 19, 2015

Revised Apr. 1, 2016

Abstract Cognitive psychology found that the cerebral cortex potential will modulate as the visual or auditory modality receive conflicting information, which can be used to explore the “stimulus-response” mechanism for the cognitive conflict control. The experiments on visual cognitive conflict is more fruitful, while little work has been conducted in the corresponding auditory ones, and inconsistent conclusions have been reported. In this study, we proposed a time feature domain of 3-stage auditory cognitive control by utilizing conflicting and non-conflicting speech signals as stimuli and analyzing and studying the corresponding electroencephalogram (EEG) signals. The paper study the feature extraction for single trial EEG sample based on cognitive control mechanism of speech conflicting. Based on the cognitive mechanism, each EEG sample for a single trial was divided into three parts. Average amplitude in the time-domain and Lempel-Ziv Complexity (LZC) was computed for the divided part of each EEG sample and the union of the feature of the three

* 国家自然科学基金项目 (61171186, 61271345)、语言语音教育部微软重点实验室开放基金项目 (HIT.KLOF.20110xx)、中央高校基本科研业务费专项基金 (HIT.NSRIF.2012047) 和全军医学科研计划面上项目 (CJN13J004) 资助

stages was used as the features of each auditory cognitive EEG sample. The study results are as follows. (1) The cognitive conflicting related complex EEG component “N1-P2&N2&Late Slow Wave (Late-SW)”, which was first found, respectively represents each stage of the auditory cognitive control. (2) A more completed auditory cognitive control process should three stages time domain feature: perceptual stage: 110~140 ms, identification stage: 260~320 ms, resolution stage: 500~700 ms. (3) The feature extraction method for a single trial auditory cognitive control EEG sample was proposed, the combination usage of Average amplitude and LZC can achieved the best accuracy (99.33%). The experimental results showed that the proposed method can effectively detect the EEG data of auditory cognitive control. Therefore, we offer an acoustic technique evaluating the ability of our brain auditory cognitive control.

引言

在声学领域, 人脑听觉认知系统的研究以及相关应用受到了越来越多的关注。人类的听觉认知系统要比目前计算机领域的自动语音识别系统复杂可靠的多, 可以帮助我们在复杂的噪声环境下进行合适的自我调节, 进行当前任务信息的保持、注意分配、反应选择, 使我们具有能够胜任某种听觉任务。这种能力指的就是“听觉认知控制”。听觉认知控制使我们能够将注意力放在当前任务相关的语音刺激属性(如: 语义、音量或方位等), 同时能够忽略或抑制其它和任务不相关的干扰刺激。因此, 听觉认知控制和社会生活息息相关。如: 在“鸡尾酒会”的环境下, 人们能够从各种嘈杂的声音中选择性地关注和识别听觉通道的输入语音信息^[1]; 当你在工作或者学习时, 你可以忽略精彩的世界杯比赛传来的欢呼声, 集中注意力到当前任务。听觉认知控制对于我们成功地工作、学习和生活有着重要的作用和意义。听觉认知控制系统是接受声音信号的重要部分, 对于其进行评价有重要实际意义。本文主要探究语音冲突状态下人脑听觉认知控制系统的工作机制, 以及如何对人脑听觉认知控制系统产生的大脑皮层的脑电信号进行自动实时地检测识别。

对人脑的听觉认知控制的研究首先要设计实验来诱发人的听觉认知控制相关的脑活动过程。目前研究人的认知控制的实验主要有 3 种: Stroop^[2], Simon^[3] 和 Flanker^[4]。其中 Simon 和 Flanker 主要用于研究视觉认知控制。目前研究人脑的视觉^[5]、听觉^[6]、情绪^[7]、空间^[8]等的认知控制主要使用 Stroop 认知控制实验来采集相关数据处理和分析。Stroop 实验原名 Stroop 色-词干扰效应, Stroop 于 1935 年发现对实验者呈现颜色词(具有词义和颜色两个维度)并令实验者做颜色命名, 相比于颜色和词义一致情况, 不一致情况下实验者的判断反应时间更长, 准确率更低, 这一干扰现象即 Stroop 色词干扰效应。目前对于人脑听觉认知控制进行研究主要

使用听觉 Stroop 实验, 可以从感知选择、反应倾向、在线信息保持以及抑制不相关信息等方面对人脑的听觉认知控制进行全面研究。该实验是在经典的视觉 Stroop 实验基础上在听觉通道的扩展, 采用在两个维度属性上相关(一致)或者不相关(不一致)的语音材料作为实验刺激材料, 实验者接受语音刺激并进行判断任务, 通过采集相关数据进行处理, 以研究听觉认知控制的脑加工机制及相关神经指标。国外的研究者 Hamers 和 Lambert 在 1972 年^[9], 以及 Cohen 和 Martin 在 1975 年^[10]得出听觉 Stroop 效应存在与视觉 Stroop 效应类似的干扰作用的结论: 不一致的听觉刺激材料导致了反应时间的延长, 准确率也跟着降低。2005 年, 国内北京师范大学心理学院发展所的李慧和陈英通过听觉 Stroop 认知控制实验得到结论, 刺激的语义对刺激的性质的干扰作用显著高于刺激的性质对刺激的语义的干扰作用^[11]。2012 年, Donohue 等在他们之前视觉 Stroop 工作的基础上进行了听觉 Stroop 实验, 证明了认知控制是跨模态的, 听觉 Stroop 任务中的认知控制的时空特性和视觉类似^[12]。对于人视觉、情绪的认知控制能力进行研究的 Stroop 实验已经有很多, 尤其是视觉认知控制的脑加工机制已经有广泛研究, 但国内外基于听觉 Stroop 实验的相关研究较少, 尤其汉语作为语音语料的研究相当少, 从而导致缺少对人脑的听觉认知控制进行研究的相关实验数据以及相关检测指标、认知规律和神经机制。听觉 Stroop 实验对研究人脑听觉认知控制的加工机制具有重要意义, 但是目前对基于听觉 Stroop 实验的人脑听觉认知控制研究仍不足以很好地解释其形成机制。只有对于听觉认知控制实验的设计和相關处理分析, 才能对人脑的听觉认知控制进行全面研究。

国内外对于脑听觉认知控制的研究主要是先通过认知控制实验获得相关数据, 然后通过数据处理方法得到相关检测指标, 最后分析得出认知规律和神经机制。目前基于认知控制实验的检测出听觉认知控制的相关指标主要有: 实验者的响(反)应时间和错误率^[13]、脑功能性神经影像得到的脑激活区^[14]、脑电

信号的时域特征结果 (主要是事件相关电位, Event Related Potential: ERP)^[12] 和脑电信号的频域特征结果^[15]。根据响应时间和错误率的检测指标对于实验者进行认知控制能力的评价是目前常用的方法。但是,如果实验者在认知控制下降或缺失的情况下进行了正确的判断,那不能正确评价实验者的认知控制能力。所以,光用行为学数据不能实时地、精确地评价实验者的认知控制能力,只能做初步的分析。目前,脑功能性神经影像得到的脑激活区的相关研究主要使用的功能磁共振成像 (fMRI),有较好的空间分辨率 (mm),但时间分辨率 (s) 较低,同时由于成本较高,对实验者和环境要求高,没有明确的认知控制相关的检测指标和神经机制,在人的认知控制能力评价上还不能广泛应用。听觉认知控制的 ERP 研究类似于视觉认知控制的 ERP 研究,关于认知控制发生的时期是注意加工的早期还是晚期的争论持续了很久。研究发现在不同实验条件下出现不同的相关听觉认知控制脑电的时域成分,分布于各个时期。由于缺少全面客观的听觉认知控制脑电的时域检测指标,无法用于人的认知控制能力的精确评价上。脑电信号的频域分析可以保留大脑神经元的非锁相信息,能更真实地得到脑认知控制相关信息,但听觉认知控制的脑电节律方面的研究还很少,并且以上关于各个脑电节律在听觉认知控制中并没有给出明确的功能,没有研究各个脑电节律如何在听觉认知控制任务中各自的贡献,尤其是高频成分的作用没有给出,从而无法建立完整的脑电节律指标,导致无法用脑电节律来衡量人的听觉认知控制能力。另外,由于已有研究的脑电信号的时频域指标的计算方法主要是叠加平均,缺少自动地、实时性地听觉认知控制检测的声学方法。

综上,关于人脑听觉认知控制系统的研究存在以下 3 方面的问题:(1) 基于语音冲突的听觉认知控制实验较少,汉语作为实验刺激的就更少,并且检测得到的大脑皮层的关于听觉认知控制的脑电信号的时域指标 (如: ERP 等) 存在不一致的问题。(2) 由于研究得到不同听觉认知控制相关的时域指标结论,从而导致不能充分地揭示在语音冲突状态下人脑听觉认知控制系统的规律和机制。(3) 对于听觉认知控制系统在大脑皮层产生的脑电信号进行处理计算,所得到的相关指标主要是基于多次试验叠加平均,缺少在语音冲突加工的实验中,对实验者单次试验获得的数据进行处理加工的方法,从而导致缺少实时地检测和识别听觉认知控制的声学方法。针对这三方面的问题,本文提出一种声学解决方法。

本文研究的主要目标是基于发现的人脑处理冲突语音的听觉认知控制的规律和机制,提出从单次试验提取认知控制相关的脑电指标,通过模式识别方法识别听觉认知控制类型,然后建立自动地、实时地对人脑听觉认知控制进行检测的方法,从而可以对人脑的听觉认知控制能力进行客观评测提供声学方法。首先,设计基于汉语语音刺激 (包括冲突和非冲突的语音刺激) 的 Stroop 实验,采集实验者的脑电波数据;然后,处理分析听觉认知控制的脑电样本数据,揭示大脑皮层相关的电位的时域指标,研究人脑进行听觉认知控制的认知规律及相应机制;进而提出基于认知规律从产生的听觉认知控制脑电样本中提取相应的认知控制特征指标的方法;最后,根据单次听觉认知控制的脑电数据样本的特征提取方法,使用支持向量机进行模式分类识别。

通过采用叠加平均方法和短时窗的统计分析方法发现了认知冲突相关的混合事件相关电位成分“N1-P2&N2&Late-SW”,表明了听觉认知控制应包括 3 个阶段。因此,一个更完整的听觉认知控制过程应包括 3 个阶段:感知阶段: 110~140 ms,识别阶段: 260~320 ms,解决阶段: 500~700 ms。基于得到的听觉认知控制的认知规律,提出基于单次试验的脑电样本进行分割的特征提取方法,联合使用平均幅度和 LZC 获取的脑电检测的指标,进而提出听觉认知控制相关的指标。实验结果证明该声学方法具有较高的识别率,可以自动地、实时地对人脑的听觉认知控制情况进行检测和识别,为客观准确地评测人脑的听觉认知控制能力提供了声学方法,具有重要的实际意义。如:通过检测汽车或飞机驾驶员的听觉认知控制,及时发现是否出现听觉通道的疲劳状态,防止意外事故发生;检测物质成瘾者在某些对其有干扰的刺激下的听觉认知控制,采取相应的戒瘾措施,减少依赖;发现听觉认知控制功能受损的人群 (老年人、中风患者等),及时采取治疗措施,延缓症状加重,改善患者生活质量。

1 听觉认知控制脑电实验

1.1 听觉认知控制

听觉认知控制是指,人们在完成听觉相关的任务中,通过听觉任务信息的维持或抑制、注意分配、听觉通道信息的感知选择、识别确认、反应决策,使我们具有能够胜任某种听觉任务的认知心理过程。神经科学的一些研究认为前扣带回 (ACC: Anterior Cingulate Cortex) 和前额皮层 (PFC: Prefrontal Cortex) 起着关键的作用,当加工冲突信息时 ACC 先激活,

负责检测或监控冲突,而 PFC 起控制调节作用^[16]。基于这种观点,Braver 提出了反馈回路模型,环境刺激产生了冲突,ACC 监测到冲突,向已经降低了控制水平的 PFC 发出唤起信号,PFC 由此加强控制,调整策略,使冲突水平降低。从而,听觉认知控制的神经机制得到了一定解释,但其认知规律还不清楚。

1.2 听觉认知控制实验的设计

听觉 Stroop 实验可以研究在语音信息冲突状态下人脑的听觉认知控制,从认知冲突(不一致或不相关)和认知一致(相关)两类语音材料听觉认知控制加工过程进行比较研究。该实验中,认知冲突情况下,需要调动人脑听觉通道的感知选择、注意、反应倾向、决策以及对于当前不相关的信息进行抑制的听觉认知控制能力。同时,通过听觉 Stroop 实验可以研究认知无关(感知阶段)和认知有关(识别阶段、执行阶段)的听觉认知控制规律。所以,通过听觉 Stroop 实验可以全面地研究人脑的听觉认知控制。

因为基于汉语语料的听觉认知控制脑电数据非常少,所以我们首先设计了基于汉语语料的听觉认知控制实验,采集相关行为学数据和脑电数据进行处理分析,得到听觉认知控制的认知规律。

共 11 名来自哈尔滨工业大学本科在读学生自愿作为实验者参与本次实验,7 名女性,4 名男性,年龄从 19 到 21 岁,平均 20.1 岁。实验者均无任何脑疾病史、身体健康、视力和听力正常、右利手。实验者在被告知实验过程后,遵从实验协助人员要求进行实验,实验后给予实验者相应报酬。所有语料均为实验室自行录制的标准汉语普通话发音,每个语料持续时间相同,均为 800 ms,一致和不一致语料如表 1 所示。大音量和小音量的差异为 40 dB,其中大音量比原始语音高 20 dB,小音量比原始语音低 20 dB。

表 1 语料清单

内容 \ 音量	音量	
	大	小
大声	一致	不一致
小声	不一致	一致

实验过程中,要求实验者根据所听到的语音材料属性(即音量的“大”还是“小”)做出判断。实验在实验者调整准备完毕后自行按 ↑ 键开始,在屏幕显示提示语“请根据声音的属性判断是大声还是小声,大声请按 ↑ 小声请按 ↓”,随机播放 4 种语料,每种所占比率均等,各 10 次,共 40 个。每个刺激呈现时间为 800 ms,两个刺激播放开始的间隔为 3300 ms,语料音量均在 60 dB 以下人耳接听声音舒适音量范围

内,要求实验者在实验中尽量快地做出判断。实验过程中屏幕中央显示的符号“+”可以帮助实验者集中注意力。由于语料的语义加工对于人脑来说是自动加工,而此实验任务要求判断音量,所以需要实验者通过认知控制来完成不一致语料的判断。

1.3 听觉认知控制实验数据采集

实验过程中,脑电信号记录设备为 NeuroScan 公司生产的 64 导 Ag/AgCl 电极帽,10/20 电极排布系统,及配套记录分析系统,进行即时记录,参考电极置于左侧乳突,滤波带通 0.01~100 Hz,采样率 1000 Hz/导,所有电极的阻抗小于 5 kΩ。Presentation 15.0 软件同步记录实验者的行为学数据,即实验者接受刺激并做出决策的按键反应时间,以及按键判断的正确与否。

2 自动听觉认知控制检测方法 & 数据处理

2.1 听觉认知控制脑电识别检测系统模型框架

图 1 显示了听觉认知控制脑电检测识别系统模型框架,主要的流程如下:

步骤 1: 设计基于判断冲突语音的听觉认知控制实验,采集实验者的脑电数据;(具体过程见 1.2, 1.3 节)。

步骤 2: 使用 ERP 方法分析得到听觉认知控制相关的时域脑电指标,进而得到听觉认知控制的机制。

步骤 3: 基于完整的听觉认知控制系统时域模型,将单次试验脑电数据作为识别样本,提取特征。每段脑电样本被分割为 3 个部分。被分割的每个阶段使用时域上的平均幅值和 Lempel-Ziv 复杂度 (LZC) 进行计算,从而联合 3 个阶段的特征作为一个听觉认知控制脑电样本的特征。

步骤 4: 优化支持向量机 (Support Vector Machine, SVM),寻找提高识别率的相关参数,训练出用于识别的模型,识别给出关于测试脑电样本的分类结果(识别是冲突语音信号的认知控制类型或非冲突语音信号的认知控制类型)。

2.2 听觉认知控制脑电实验数据的预处理方法

由于,眨眼、眼动、心电等信号对听觉认知控制脑电信号会产生影响,所以首先对采集得到的听觉认知控制脑电信号进行预处理。对一个实验者一个脑电电极采集得到的听觉脑电 ($\mathbf{X}(n)$: $\mathbf{X}_1(n), \mathbf{X}_2(n), \mathbf{X}_3(n), \dots, \mathbf{X}_N(n)$)(其中, $\mathbf{X}_i(n)$ 为第 i 次试验得到的脑电数据) 进行预处理的过程主要包括: 去除眼电、脑电分段、基线校正、自适应去除伪迹。

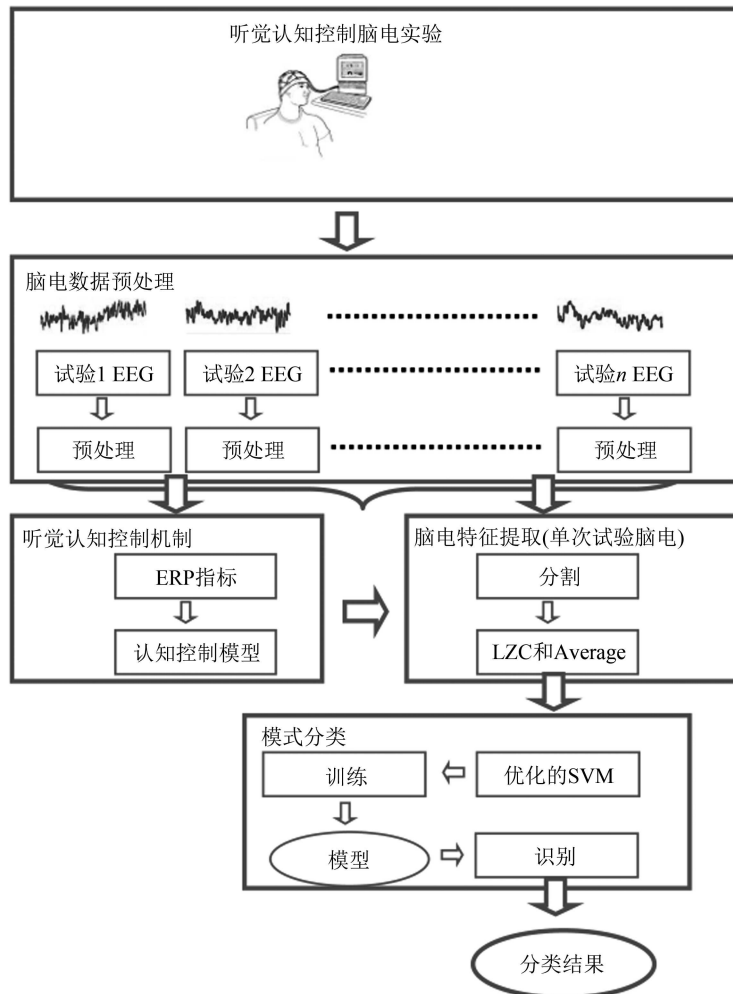


图 1 听觉认知控制脑电识别检测系统模型框架

2.2.1 去除眼电

采用平均伪迹线性回归分析方法从原始脑电信号 $X(n)$ 中删除 $\mathbf{VEOG}$ (垂直眼电) 得到修正后的脑电信号^[17], 如式 (1):

$$\mathbf{X}_c(n) = \mathbf{X}(n) - \beta \times \mathbf{VEOG}, \quad (1)$$

其中, $\mathbf{X}_c(n)$ 是去除眼电后的脑电信号, β 为回归因子, 也即眼电电极和头皮电极之间的传导系数。

为了减少眨眼活动和连续脑电活动之间的相关性, $\mathbf{VEOG}$ 和脑电电位在眨眼上进行平均 (最小眨眼数为 20, 持续时间为 400 ms)。眼电电极超过 $\mathbf{VEOG}$ 最大值的 10% 的电位被识别为眼电脉冲, 然后对识别的眼电脉冲进行平均, 得到平均 $\mathbf{VEOG}$ 。各个电极的回归因子 β 的估计分别由每个电极的平均脑电电位和平均 $\mathbf{VEOG}$ 来进行计算, 如式 (2):

$$\beta = \frac{\text{cov}(\mathbf{X}(n), \mathbf{VEOG})}{\text{var}(\mathbf{VEOG})}, \quad (2)$$

其中, cov 为协方差计算, var 为方差计算。

2.2.2 脑电分段

将修正后的连续脑电数据 $\mathbf{X}_c(n)$ 进行分段, 每段为语音信号刺激播放前 200 ms 和语音信号刺激播放后 823 ms, 得到分段后的脑电: $\mathbf{X}_{c1e}(n)$, $\mathbf{X}_{c2e}(n)$, $\mathbf{X}_{c3e}(n)$, $\dots$, $\mathbf{X}_{cNe}(n)$ 。

2.2.3 基线校正

基线校正可以消除脑电相对于基线的偏离。基线校正的刺激前的时间长度为 200 ms, 作为基础值。将脑电电位减去该基础值, 得到: $\mathbf{X}_{c1eb}(n)$, $\mathbf{X}_{c2eb}(n)$, $\mathbf{X}_{c3eb}(n)$, $\dots$, $\mathbf{X}_{cNeb}(n)$, 从而删除直流偏移。

2.2.4 去除伪迹

去除伪迹的目的主要是将分段、基线校正后的脑电数据中存在幅度较高的伪迹删除 (如: 肌电、高波幅的慢电位等)。为保证足够的试验次数用于后续分析, 提出自适应伪迹去除。去除的阈值从 70 μV 开始, 每次递增 5 μV , 直到去除伪迹的比率小于 20% 停止, 阈值终止值为 150 μV , 得到 $\mathbf{X}_{c1eba}(n)$, $\mathbf{X}_{c2eba}(n)$,

$X_{c3eba}(n), \dots, X_{cNeba}(n)$ 。

2.3 听觉认知控制脑电实验数据的处理方法

2.3.1 计算事件相关电位 (ERP)

将预处理之后得到的每个实验者每个脑电电极的脑电数据采用ERP方法进行时域处理。在得到原始脑电数据预处理后的脑电数据 ($X_{c1eba}(n), X_{c2eba}(n), X_{c3eba}(n), \dots, X_{cNeba}(n)$) 基础上, 应用式 (3) 计算每一种语音刺激条件下的 ERP。

$$ERP(n) = \frac{\sum_{i=1}^N X_{cieba}(n)}{N}, \quad (3)$$

2.3.2 转换参考

为了避免结果分布失真, 将左右乳突的平均作为叠加平均后的脑电数据的参考。设左侧乳突的原始电位值为 L , 电极 X 的原始电位值为 X , 记录得到电位值为 $X-L$, 设为 A 。设右侧乳突的原始电位值为 R , 记录得到电位值为 $R-L$, 设为 R' 。转换为左右乳突的平均数为参考后, X 位置的信号幅值为 X' :

$$X' = X - \frac{L+R}{2} = A + L - \frac{L+R}{2} = A - \frac{R'}{2}. \quad (4)$$

2.3.3 ERP 滤波

为了得到光滑的 ERP 波形和消除 50 Hz 的市电干扰, 将转换参考后的 ERP 脑电数据进行 30 Hz 的低通滤波, 相位漂移设定为零。

2.4 听觉认知控制脑电实验数据的统计分析

所得行为学数据用 IBM SPSS Statistics 19.0 软件进行统计分析处理。对行为学数据反应时间和错误率均求取总平均, 并进行重复测量方差分析 (ANOVA)。

对 530 ms 之前的各实验者的 ERP 数据的统计分析基于的是被试内因子模型, 进行三因素重复测量 ANOVA: 脑区侧性 (左、中、右) × 脑区前后位置 (前、中、后) × 语料类型 (一致、不一致)。对 50 ms 至 530 ms 之间的数据加窗分段, 为了能精确分析早期的 ERP 成分, 采用更短的时间窗, 窗长 30 ms。选取跟听觉认知控制密切相关的脑区 (额区 - 中央区 - 顶区中央线及其周边) 所对应的 9 个电极 (在本实验所得 N120 和 N300 最大值均在此中线上): F1, FC, C1, FZ, FCZ, CZ, F2, FC2, C2。自由度 ε 进行自适应调节, 如果 $\varepsilon < 0.75$, 采用 Greenhouse-Geisser 方法修正; 如果 $1 > \varepsilon > 0.75$, 采用 Huynh-Feldt 方法修正。如果一个因子包含两个水平有主效应, 使用 Bonferroni 测验进行 Post-hoc 分析。如果多个因子存在交互效

应, 执行简单效应分析。

2.5 听觉认知控制脑电数据特征提取方法

因为脑电信号的时域分析方法包含了一个 EEG 信号的所有信息, 所以使用计算一个脑电序列的平均幅度的方法。该方法有很高的计算效率。式 (5) 给出计算一个给定脑电序 $S(S(1), S(2), \dots, S(N))$ 的平均幅度 (Meanamplitude) 的方法。

$$\text{Meanamplitude} = \frac{1}{N} \sum_{i=1}^N S(i), \quad (5)$$

其中, $S(i)$ 表示语音刺激播放后, 第 i ms 的脑电电位幅值。

Lempel 和 Ziv 于 1976 年首次提出 LZC^[18]。它是一种反应新模式在时间序列里出现的新模式的非线性方法。LZC 的值越大代表新模式出线得频率越高, 表明有更复杂的脑动态行为。下面给出对于输入的一个脑电时序序列 $S(S(1), S(2), \dots, S(N))$, LZC 的基本计算过程如下:

(1) 二值化序列 S , 设定阈值, 大于阈值设定为 1, 小于阈值设定为 0。

(2) 初始化, $C=1, S=S(1), Q=[], SQv=[]$ 。

(3) 把 S 序列中的下一个符号级联到 Q , SQ 为级联 S 和 Q , SQv 表示从 SQ 中去掉最后一个符号。

(4) 如果 $Q \in SQv, C=C+1, S=SQ$, 清空 Q 。

重复过程 (3), (4) 直到 Q 取到待求序列的最后一个。

(5) 使用式 (6) 计算 LZC。

$$LZC = \frac{\log_2 N * C}{N}. \quad (6)$$

结合计算脑电平均幅度和 LZC 的方法, 给出本文提出的针对听觉认知控制听觉单次试验的脑电样本的特征的方法。输入的待处理数据为使用 2.2 节预处理之后的脑电数据, 输出为单次试验脑电样本的特征向量。具体的计算过程如下:

(1) 访问每个实验者的脑电数据的单次试验的脑电数据。

(2) 获得 18 个电极 (F1, FZ, F2, FC1, FCZ, FC2, C1, CZ, C2, CP1, CPZ, CP2, P1, P2, PO3, POZ, PO4) 的脑电数据。

(3) 将每个电极的脑电数据分成 4 个时间段 (感知阶段: 110~140 ms, 识别阶段: 260~290 ms, 290~320 ms, 执行阶段: 500~700 ms)。

(4) 计算每个时间段的平均幅度和 LZC 及每次试验整段脑电数据的 LZC。

(5) 生成单次试验脑电样本的特征向量。

重复上述计算过程, 遍历每个实验者, 每次试验的脑电数据, 依次将每次试验的特征向量添加到特征矩阵, 用于模式识别。

2.6 听觉认知控制脑电的检测识别方法

采用优化的 SVM 进行模式识别。SVM 的主要思想是找到一个最优的超平面, 使得两类样本之间距离最大。假设给一个线性可分的训练样本集 $\{(\mathbf{x}_i, z_i) | \mathbf{x}_i \in R^d, z_i \in \{-1, 1\}\}$, $i = 1, 2, \dots, n$ 。对于非线性问题, 通过非线性映射函数 $\Phi(\mathbf{x}_i)$, 样本被从低维输入空间映射到高维特征空间, 从而使训练样本可分。为了避免空间变换引起样本维数增加的问题, 满足 Mercer 条件的核函数, 直接在低维空间计算样本在高维空间的内积值, 即: $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)$ 。最终, SVM 的分类函数如下式 (7) 所示。

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^n \alpha_i^* z_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right), \quad (7)$$

其中, α_i 为拉格朗日乘子, 可以由如式 (8) 的凸最优化问题求解。

$$\begin{aligned} \min_{\alpha \in R^n} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n z_i z_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{j=1}^n \alpha_j, \\ \text{s.t.} & \sum_{i=1}^n z_i \alpha_i = 0, 0 \leq \alpha_i \leq c, i = 1, \dots, n. \end{aligned} \quad (8)$$

径向基函数 (RBF, Radial Basis Function) 作为核函数, 见式 (9)。通过优化训练模型来优化 SVM, 通过使用 K 折叠交叉验证 (K-CV, K-fold Cross Validation) 算法来选择 SVM 的优化参数。使用 RBF 的

SVM 模型有两个可调参数 (g : RBF 函数中的参数; c : 惩罚因子, 见式 (8))。

$$K(\mathbf{x}, \mathbf{y}) = e^{(-g|\mathbf{x}-\mathbf{y}|^2)}. \quad (9)$$

3 自动听觉认知控制检测相关实验结果及分析

3.1 听觉认知控制脑电实验结果及分析

3.1.1 听觉认知控制的行为学结果及分析

图 2 显示了实验者对于语音材料听觉认知控制反应的行为学结果。可以看出, 实验者对于不一致语音材料平均反应时间 (790.7 ms) 显著慢于一致语音材料 (688.9 ms) ($F(1, 12) = 32.62, p < 0.05$), 平均错误率则表现为不一致语料 (6.57%) 显著高于一致语料 (0%) ($F(1, 11) = 15.48, p < 0.05$)。说明实验者在完成不一致任务时需要调用更多的脑听觉认知控制资源来完成当前任务, 并抑制不相关的信息。可以通过行为学的指标对于人脑听觉认知控制能力进行初步评价。

3.1.2 听觉认知控制脑电的时域表现结果及分析

11 个实验者有相似的波形图, 主要表现为 4 个 ERP 脑电成分: N1 (峰值在 150 ms 附近), P2 (峰值在 220 ms 附近), N2 (峰值在 290 ms 附近) 和 Late-SW (晚期慢波, 范围在 500~823 ms)。体现了完整的听觉认知控制的加工过程。对所有实验者个体的 ERP 求总平均, 得到如图 3 所示的 ERP 波形。

方差分析结果显示 (详见表 2), N1 和 N2 在人

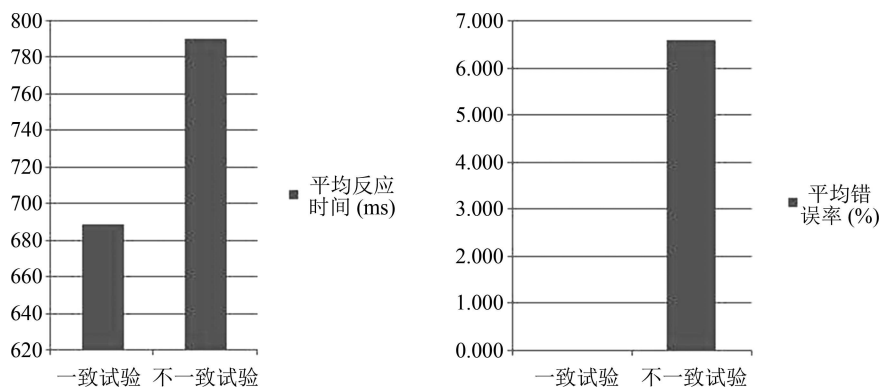


图 2 行为学结果

表 2 三因素重复测量方差分析结果

时段 (ms)	语料类型	时段 (ms)	语料类型 × 侧性	时段 (ms)	语料类型 × 前后位置	时段 (ms)	语料类型 × 侧性 × 前后位置
110~140	$F = 6.65, p = 0.027$	320~350	$F = 3.61, p = 0.046$	440~470	$F = 6.54, p = 0.007$	80~110	$F = 4.28, p = 0.006$
260~290	$F = 7.06, p = 0.024$	440~470	$F = 5.15, p = 0.042$	470~500	$F = 5.55, p = 0.032$		
290~320	$F = 4.97, p = 0.050$	470~500	$F = 6.17, p = 0.021$				
		500~530	$F = 6.85, p = 0.005$				

脑听觉认知控制的加工过程中同时存在。在 110~140 ms ($F = 6.652, p = 0.027$) 的时间段, 不一致语音刺激引起的 N1 的电位幅值比一致刺激要更负; 在 260~290 ms ($F = 7.056, p = 0.024$) 以及 290~320 ms ($F = 4.971, p = 0.050$) 的时间段, 不一致语音刺激引起的 N2 的电位幅值比一致刺激要更负。

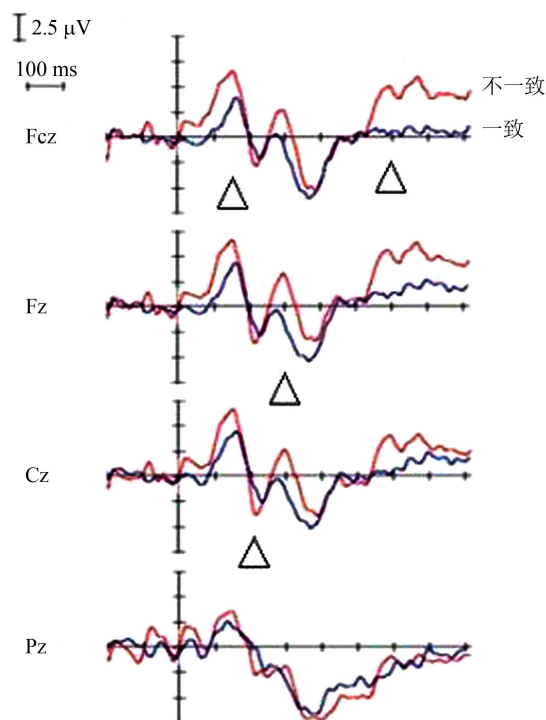


图 3 4 个脑中线临近电极的各实验者的总平均 ERP

听觉认知控制的脑电电位的相关研究中, Ninc (300 ms, 大约在 N2 的位置) 和 Late-SW 两个差异成分在 Donohue 等人^[12]的研究中已经分离出来, 而 N1 (110~140 ms) 和 P2 (200~250 ms) 左右出现的差异成分是前人所未观察到的。Henkin 等人的研究发现了听觉认知控制相关的 N1 成分, 但却没有指出 N2 成分的存在。

一致减去不一致的 ERP 成分所对应的脑地形图如图 4 所示。这些 ERP 成分 (N1, P2, N2, Late-SW) 主要分布在前额区和中央区。从脑电地形图的情况来看, 不一致减一致所获得的差异脑电成分分布的位置基本上在前额区和中央区, 因此这几个成分的确是听觉认知控制相关的任务成分。听觉认知控制相关的差异成分的偏负性体现了对干扰信号的加工抑制。

3.2 听觉认知控制脑电样本的检测识别结果及分析

3.2.1 实验 1: 使用不同特征

针对听觉认知控制脑电数据特征提取方法所构建的特征矩阵进行 5 折叠交叉验证, 用于选择优化模式识别方法的参数, SVM 的 g 和 c 的寻优范围为 $2^{-10} \sim 2^{10}$, 指数部分变化步长为 2。70% 的脑电样本被随机选择作为训练样本, 另外 30% 作为测试样本。相比较于直接使用样本特征, 正规化的样本特征可以改进识别率。试验结果为 100 次试验的平均值。表 3

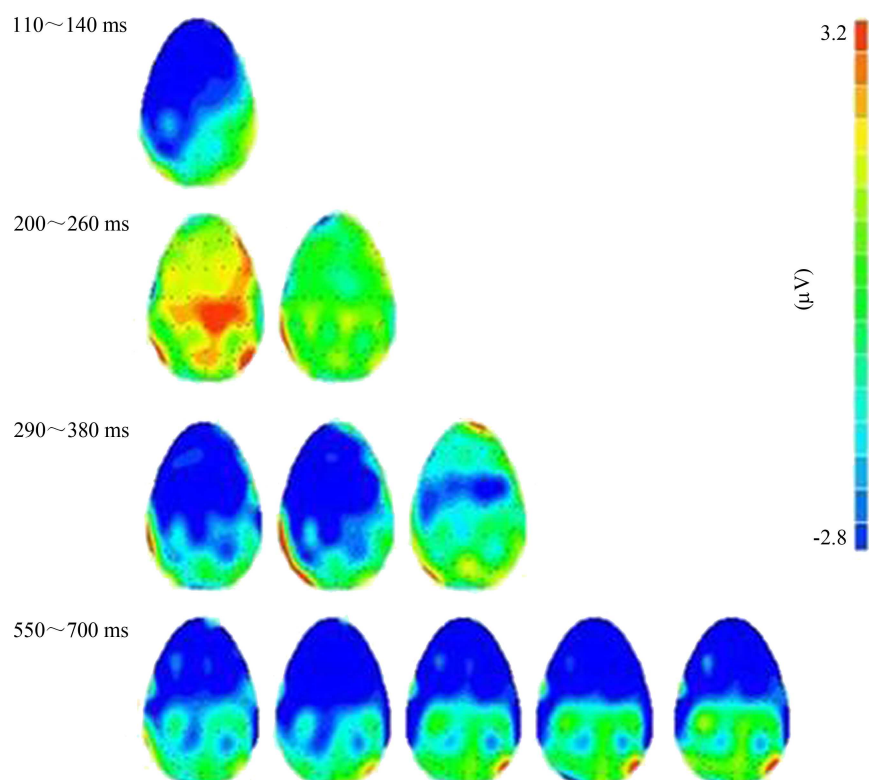


图 4 4 个 ERP 成分 (N1, P2, N2, Late-SW) 所在的 30 ms 的时间窗内的平均幅值的脑电地形图 (不一致减一致)

显示了使用不同的正规化脑电特征的识别率，可以看出联合使用平均幅度和 LZC 可以获得最好的识别率。

表 3 使用不同特征提取方法的识别率

特征	正确率	F-value ^a
Average	87.21%	0.81
LZC	70.65%	0.66
Average+LZC	99.33%	0.98

a: 听觉认知控制样本的 F-value

基于提出的三阶段的听觉认知控制模型提取特征，可以利用听觉认知控制规律。而平均幅度可以较好地描述每个脑电时间段各个时间点脑电幅值的局部变化情况。而 LZC 可以从每个脑电时间段整体上描述出线新模式的情况，定量描述信号曲线变化的复杂性。实验结果表明平均幅度提取特征的方法要好于 LZC，原因在于平均幅度更体现了听觉认知控制过程的 ERP 规律。而联合使用平均幅度和 LZC 特征，将人脑听觉认知控制过程的局部特征和全局特征综合考虑，形成互补，从而获得了最高的听觉认知控制脑电检测识别率。

3.2.2 实验 2: 使用不同听觉脑电时间段作为特征

为了进一步验证平均幅度和 LZC 特征对于脑电的听觉认知控制的表征，针对不同时间段提取的平均幅度和 LZC 脑电特征作为一次脑电试验样本的特征进行模式识别。针对每个时间段进行特征提取和模式识别的过程同实验 1。表 4 显示了使用不同时间段提取的特征的识别率。

表 4 使用不同时间段提取的特征的识别率

时间段	正确率	F-value ^a
110~140 ms	80.21%	0.74
260~320 ms	79.65%	0.73
500~700 ms	81.33%	0.76

a: 听觉认知控制样本的 F-value

表 3 的识别率结果表明使用不同时间段提取的特征的识别率并没有显著差异，都可以较好地完成识别任务。该结果说明平均幅度和 LZC 脑电特征能够表征人脑听觉认知控制系统对于语音冲突信息加工的差异（一致 vs. 不一致），即在感知阶段、识别阶段以及解决阶段存在差异，与 ERP 指标有类似的结论。同时，该结果也说明针对单次听觉认知控制的脑电样本所提取的平均幅度和 LZC 特征能够很好地表征听觉认知控制的每个阶段。

4 讨论

4.1 听觉认知控制的规律及模型

4.1.1 听觉认知控制的规律

据我们的研究工作所知，听觉认知控制只有两个使用听觉认知控制实验的 ERP 研究同时发现了认知控制的早期阶段和晚期阶段的成分，从而提供了很少的证据支持跨模态的听觉认知控制的“冲突探测-冲突解决”机制。Donohue 和 Liotti^[12] 在他们的听觉认知控制的研究中确认了“冲突探测-冲突解决”机制。他们发现了，峰值在 300 ms 的早期 ERP 成分 (Ninc, 200~500 ms)，之后是 Late-SP (Late Sustained Positivity, 晚期持续的正波，从 500~800 ms)，同时，提出这两个 ERP 成分分别对应之前视觉认知控制的研究中发现的 N450 和 Late-SP。通过序列分析表明晚期阶段主要参与认知控制的适应。另外一个由 Lew 等人^[19] 开展的更早的研究证明了听觉认知控制干扰效应存在于感知过程 (N1 成分) 和后感知 (反应) 过程 (晚期慢波成分)。遗憾的是，其他听觉认知控制的研究并没有发现完整的冲突探测-解决机制。例如，Buzzel 等人的研究^[17] 证明了 Ninc 的幅度是由认知方式的个体差异可预见地调节，并且确认了 Ninc 指示了由听觉空间 Stroop 任务引起的听觉 Stroop 冲突。Henkin 等人^[13] 发现早期的 N1 效应是通过认知控制调节，然后，他们的研究中没有发现显著的 SP (持续的正波)。

本研究的听觉认知控制任务中，加工不一致和一致刺激引发的 N2 (290 ms 左右) 和 Late-SW (范围在 500~823 ms) 的差异在 Donohue 等人的研究^[12] 中已经发现。而加工不一致和一致刺激引发的 N1 所引发的差异是 Donohue 等人所未观察到的。本研究工作同时发现在听觉认知控制任务中显著的 N1 和 N2，所以“冲突探测-冲突解决”机制中的冲突检测环节在听觉 Stroop 效应中的具体情况不仅由 Donohue 等发现的 Ninc (200~500 ms，大约在 N2 的位置) 体现，也由 Henkin 和 Lew 等认为的 N1 来体现^[13,19]，两者均参与了冲突探测。N1, P2, N2 三个差异波均在额-中央区最大，而这个区域一般认为与听觉认知控制最密切相关，均与识别加工相关，前两者为知觉成分，后者为认知成分，它们可能都参与了听觉认知冲突的探测识别，并且使得冲突探测呈现出冲突感知和冲突识别两个阶段，这就解释了为什么 Henkin 和 Yael 等发现 N1 成分是冲突探测，而 Donohue 等发现 Ninc (200~500 ms) 是冲突探测。

4.1.2 听觉认知控制的模型

N1 是知觉唤醒成分。Donohue 等未关注知觉唤醒成分在听觉认知控制中的作用。Henkin 和 Yael 等的听觉 Stroop 研究认为 N1 家族成分在 Stroop 效应中的作用是：作为冲突信息到达初级听觉皮层的信号参与到“冲突探测-解决机制”中。而在本实验的关注属性任务中，N1 的显著差异性（不一致对一致）正说明其参与了冲突探测，体现了早期知觉探测也可以进行选择性的注意的调控。

P2 是知觉成分，其最大值位于顶区（头皮后部）。在视觉认知控制的加工（如：面孔、文字等）的 ERP 研究中^[20-21]，P2 被认为是体现与视觉信息的早期语义加工有关的成分，那么可将之视为和听觉认知控制早期语义加工相关成分。

N2 是认知成分，在实验结果中稳定存在，且位于额区和中央区，说明该成分体现的偏负抑制性正是听觉认知控制效应对干扰信息的抑制加工的关键，Donohue 等人也认为正是这一成分体现了对冲突的探测。所以，该成分体现了冲突探测的识别阶段。

Late-SW 是反应决策成分，分布于额区和中央区，之前 Donohue 和 Henkin 的研究一致认为该成分反应了冲突解决。所以，该成分反应了冲突解决阶段的反应决策、动作执行。

上述 Stroop 效应中发现的完整的 ERP 生物指标证明了听觉认知控制应该具有一个相比于“冲突探测-冲突解决”更复杂的时域过程。本文提出完整的听觉认知控制的时域模型应包含“感知阶段-识别阶段-解决阶段”。基于得到的 ERP 实验结果以及行为学反应时间为 700 ms 左右，所发现的混合 ERP 成分“N1-P2&N2&Late-SW”分别体现了“感知-识别-解决”的听觉认知控制的 3 个阶段，这 3 个阶段构成一个更完整的听觉认知控制过程（如图 5 所示）。感知阶段（110~140 ms）：N1 和 P2 联合体现了对当前任务早期的选择注意；识别阶段（260~320 ms）：N2 体现了在线工作状态，以及对当前任务目标的分类

识别；解决阶段（500~700 ms）：Late-SW 反应了对当前任务的决策、反应，并通过动作执行完成认知控制任务。同时，听觉认知冲突控制系统在 3 个阶段都要对任务不相关信息进行抑制。本文所提出的听觉认知控制的模型较之前提出的“冲突探测-冲突解决”的模型更加细化、完整，尤其是对各早期 ERP 成分（N1, P2, N2）在冲突探测阶段的具体功能。

4.2 自动听觉认知控制的检测识别方法

提出了基于人脑听觉认知控制规律，对单次脑电数据进行听觉认知控制相关特征提取的方法，从而得到基于认知规律的听觉认知控制的自动检测识别的方法。为自动的人脑听觉认知控制的评价提供了手段。通过检测识别结果可以看出，提出的方法能够有效从听觉脑电数据中识别出听觉认知控制样本。同时，该检测识别结果也说明，不用基于叠加平均原理的 ERP 方法，针对单次听觉认知控制的脑电样本所提取的平均幅度和 LZC 特征也能很好地表征听觉认知控制的每个阶段。相对于其他听觉认知控制评价的方法本文提出的方法具有的特点：

(1) 成本相对较低，受环境影响小。

(2) 不需要叠加平均计算，可以实时地对采集的单次试验脑电数据进行检测识别。

(3) 实验者听觉脑电数据混合的情况下，也具有较高的检测识别率。

与传统的检测评价方法相比，比如通过错误率的高低来评价实验者的认知控制水平。错误率为实验者在一次实验会话（Session）中，错误做出判断的试验（Trial）次数占试验总数的比例。实验者对于不一致刺激的正确反应相对于一致刺激会降低，从而导致错误率降低。如果实验者在认知控制下降的情况下进行了正确的判断，那不能正确评价实验者认知控制能力。所以，光用行为学数据不能实时地、精确地评价实验者的认知控制能力，只能做初步的定性分析。而听觉脑电数据最能准确地反应实验者每次试验的听觉认知控制状态。根据错误率的评价方

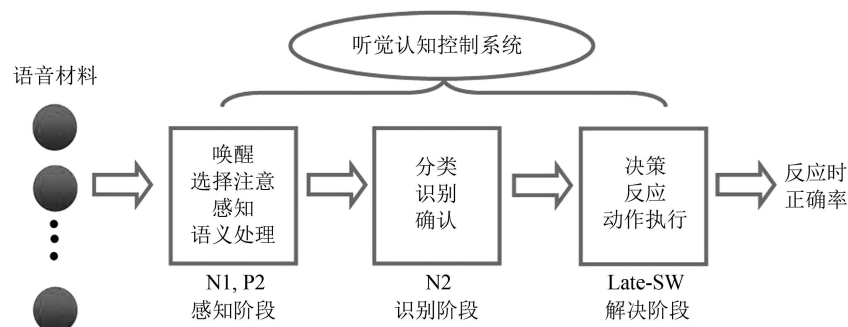


图 5 听觉认知冲突控制系统时域特征模型

法, 定义出听觉认知控制率 (ACCR: Auditory Cognitive Control Rate): 实验者在一次基于语音判断的听觉实验会话中, 认知控制类型的脑电次数占试验总数的比例。该评价指标相比于错误率能够更客观地对人脑听觉认知控制能力进行评价。

5 结论

针对自动地对人脑听觉认知控制检测的问题, 本文提出了基于人脑听觉认知冲突时域模型的听觉认知控制特征提取的声学方法。首先, 研究分析了认知控制对于人的重要性, 并提出如何自动实时地对听觉认知控制进行检测的新问题。其次, 设计了基于汉语语料的听觉认知控制脑电实验; 根据得到的 ERP 波形、行为学数据以及统计分析结果构建了听觉认知冲突控制的时域特征模型, 包括 3 个阶段: 感知阶段 (110~140 ms)、识别阶段 (260~320 ms) 以及解决阶段 (500~700 ms)。之后, 基于听觉认知控制的认知机制, 提出对待识别单次试验脑电样本进行听觉认知控制相关的特征提取方法。根据听觉认知控制的认知规律进行分段, 并联合反映 ERP 幅值情况的平均幅度和反映时间序列出线新模式情况的非线性动力参数 LZC 提出脑电特征的方法。最后, 使用优化的 SVM 进行听觉认知控制的识别。

本文提出的针对单次脑电的特征提取的声学方法可以从听觉认知脑电数据中检测识别出听觉认知控制相关脑电, 从而使用本文提出的听觉认知控制率的指标可以客观准确地对人的听觉认知控制能力进行评价, 具有较高地实用价值。本文的研究方法可以推广到视觉、情绪等方面的脑认知控制的检测评价上。提出的听觉认知控制时域特征模型对于复杂环境下的自动语音识别系统的改进也有借鉴意义。下一步的工作将研究听觉认知控制过程频域的认知规律, 挖掘相关脑波振荡的指标。

参 考 文 献

- Haykin S, Chen Z. The cocktail party problem, *Neural Comput.*, 2005; **17**(9): 1875—1902
- Stroop J R. Studies of interference in serial verbal reactions. *J. Exp. Psychol. Anim. B.*, 1935; **18**(6): 643—662
- Simon J R, Berbaum K. *Acta Psychol. (Amst.)*, 1990; **73**(2): 159—170
- Eriksen B A, Eriksen C W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 1974; **16**(1): 143—149
- Brown T L, Gore C L, Carr T H. Visual attention and word recognition in stroop color naming: is word recognition “automatic”? *J. Exp. Psychol. Gen.*, 2002; **131**(2): 220—240
- Morgan A L R, Brandt J F. An auditory stroop effect for pitch, loudness, and time. *Brain Lang.*, 1989; **36**(4): 592—603
- Williams J M, Mathews A, MacLeod C. The emotional Stroop task and psychopathology. *Psychol. Bull.*, 1996; **120**(1): 3—24
- Wuhr P. A Stroop effect for spatial orientation. *The Journal of General Psychology*, 2007; **134**(3): 285—294
- Hamers J F, Lambert W E. Bilingual interdependencies in auditory perception. *Journal of Verbal Learning and Verbal Behavior*, 1972; **11**(3): 303—310
- Cohen G, Martin M. Hemisphere differences in an auditory Stroop test, *Perception and Psychophysics*, 1975; **17**(1): 79—83
- 李慧, 陈英和. 听觉通道 Stroop 效应的实验研究. 第十届全国心理学学术大会, 北京, 2005
- Donohue S E, Liotti M, Perez R, Woldorff M G. Is conflict monitoring supramodal? Spatiotemporal dynamics of cognitive control processes in an auditory Stroop task. *Cogn. Affect. Behav. Ne.*, 2012; **12**(1): 1—15
- Henkin Y, Yaar-Soffer Y, Gilat S, Muchnik C. Auditory Conflict Processing: Behavioral and Electrophysiologic Manifestations of the Stroop Effect. *J. Am. Acad. Audiol.*, 2010; **21**(7): 474—486
- Schulze K, Mueller K, Koelsch S. Auditory stroop and absolute pitch: An fMRI study. *Hum. Brain Mapp.*, 2013; **34**(7): 1579—1590
- Oehrns C R, Hanslmayr S, Fell J, Deuker L, Kremers N A, Do Lam A T, Elger C E, Axmacher N. Neural communication patterns underlying conflict detection, resolution, and adaptation. *J. Neurosci.*, 2014; **34**(31): 10438—10452
- Botvinick M M, Braver T S, Barch D M, Carter C S, Cohen J D. Conflict monitoring and cognitive control. *Psychol. Rev.*, 2001; **108**(3): 624—652
- Buzzell G A, Roberts D M, Baldwin C L, McDonald C G. An electrophysiological correlate of conflict processing in an auditory spatial Stroop task: the effect of individual differences in navigational style. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 2013; **90**(2): 265—271
- Lempel A, Ziv J. On the Complexity of Finite Sequences. *Information Theory, IEEE Transactions on*, 1976; **22**(1): 75—81
- Lew H, Chmiel R, Jerger J, Pomerantz J R, Jerger S. Electrophysiologic indices of Stroop and Garner interference reveal linguistic influences on auditory and visual processing. *J. Am. Acad. Audiol.*, 1997; **8**(2): 104—118
- Zhao L, Li J. Visual mismatch negativity elicited by facial expressions under non-attentional condition. *Neuroscience Letters*, 2006; **410**(2): 126—131
- Peressotti F, Pesciarelli F, Mulatti C, Dell’Acqua R. Event-related potential evidence for two functionally dissociable sources of semantic effects in the attentional blink. *Plos One*, 2012; **7**(11): e49099