



基于语料语义的脑电反应一致性研究

摘要

通过测量语音相关的脑电信号变化来研究人脑对于标准语料的反应趋势。有16名被测者分别聆听了120条标准语音,每条8 s,语音之间有1~2 s的间隔且随机播放。提取被测者在聆听过程中的脑电信号,对1~40 Hz频段的脑电信号进行预处理并将其与语音信号进行分析对比。结果显示,被测者听到相同标准语音时有相同的脑电反应趋势。通过相锁值方法对脑电信号和语音信号进行相位差分析,证明了脑电信号与语音质量之间的功能连接性,并且脑电信号区分语音质量的正确率达到99.62%。

关键词

脑电信号;语音质量;感知评估;自然刺激

中图分类号 TN912.34;TN911.7

文献标志码 A

收稿日期 2024-07-05

资助项目 国家重点研发计划“变革性技术关键科学问题”重点专项(2018YFA0701601)

作者简介

宋奇蔚,女,博士生,高级工程师,研究方向为移动通信网络、业务质量评估。sqw19@mails.tsinghua.edu.cn

葛宁(通信作者),男,博士,教授,研究方向为通信、网络相关理论及应用研究。gening@tsinghua.edu.cn

1 清华大学 电子工程系,北京,100084

0 引言

得益于高速发展的移动通信技术,语音通话可以提供一系列的服务,包括高清语音通话、多方通话、视频呼叫、实时字幕和语音助手等。而这些服务也涉及众多不同的厂商,如网络运营商、因特网服务供货商、通信软件提供商等。因此,如何在各种服务下确保语音通话质量就变得尤其重要。

传统的语音质量主观评价方法以MOS(Mean Opinion Score,平均意见分)评价为主,该方法邀请用户对语音服务进行打分,从而获得用户对语音质量和满意度的主观评分。但是,传统评价方法容易受到外界因素的影响,使用成本高昂并且实现周期长。因此,国际电信联盟(International Telecommunication Union, ITU)下设的国际无线电咨询委员会(International Radio Consultative Committee, CCIR)提出了基于客观语音质量评估的标准ITU-T P.862 PESQ(Perceptual Evaluation of Speech Quality, 语音质量感知评估)和ITU-T P.863 POLQA(Perceptual Objective Listening Quality Analysis, 感知客观听力质量评估)。以上两个标准都是基于感知模型的客观语音质量评估方法,主要不同在于认知模型和算法细节。

目前用于主观评价所构建的感知模型尚不能完全模拟人的听觉感知,有可能存在听觉感知好但评分低的情况。因此,研究者们正在寻求其他客观感知评价的方法。文献[1]提出使用脑电信号来评估语音质量,证明了使用基于脑电图描述神经元在语音质量感知方面差异的测量方法可以客观测量受试者的听觉感知,并且不受自我意识干扰的影响。文献[2]对脑电信号特征进行分析,结果表明,P300分量与语音质量有较为明显的相关性。为了验证语音语料和脑电反应具有一致性,本文采用自然刺激的方式,让被测者接受标准语料的刺激,同时记录他们的脑电数据,通过脑电数据分析大脑活动模式,评估被试者对语音的响应和感知情况。此外,本文还分析了脑电信号与语音质量之间的关系,通过深度学习网络验证脑电信号区分5种语音质量的正确率达到99.62%。

1 脑电信号和语音感知评价

1.1 脑电信号简介

EEG(Electroencephalogram, 脑电信号)是一种记录大脑内神经元

电活动的生理信号.通过在头皮上放置电极,获取当前状态下的脑电信号,以便对大脑的功能活动进行研究.脑电信号研究在神经科学、认知心理学和临床医学等学科领域中被广泛应用.脑电信号具有高时间分辨率和非侵入性等特点,多年研究积累形成了神经振荡、神经同步和事件相关电位等理论框架.通过分析脑电信号与行为表现之间的关联性,揭示了大脑活动与认知行为(如注意、记忆、语言和情绪等)之间的关系.上述神经振荡、神经同步和事件相关电位等理论为基于脑电信号的语音质量感知评价方法提供了新的研究方向.

1.2 语音感知评测

传统的语音质量评测一般采用 MOS 值表示,但该分值是用户的主观评测,不一定能真实反映语音的质量.因此,ITU 提出了基于客观感知的评测方法.语音服务感知评估是从人的听觉机制角度对语音内容的理解以及用户的满意程度来评测语音服务质量的.目前的主流方法是 ITU-T P.863 标准中的 POLQA 算法,其核心是通过模拟听感,用客观指标表述语音感知,然后通过客观性指标和补偿性指标计算出 POLQA 值,经由多项式映射为 MOS-LQO (Listening Quality Overall, 总体听力质量)值,从而达到评测目的^[3].

1.3 脑电信号分析方法

在传统的认知科学实验范式中,研究者通常使用相对简单的声音刺激,如纯音或词汇来研究听觉注意的现象.然而,简单的声音刺激并不能完全模拟日常生活中复杂多变的听觉场景,因此该范式存在一定的局限性,在生态效度方面有待提高.

近年来,越来越多的认知科学研究开始尝试采用连续语音作为声音刺激,并将语音理解作为主要的听觉任务.这种新的研究方法旨在更真实地模拟日常生活中的听觉场景,从而提高研究的生态效度.通过这种方法,研究者能够更深入地探索大脑在复杂、高生态效度的听觉场景中的听觉注意机制.这种范式的转变有助于更好地理解大脑在现实生活中的听觉处理过程,为相关领域的研究提供更具有实际意义和价值的科学依据.

研究发现,当人们聆听语音时,大脑的 delta 频段(不大于 4 Hz)和 theta 频段((4,7] Hz)的神经活动会与语音进行同步跟踪^[4-5].其中,大脑的低频神经活动会同步跟踪语音中的低层级声学特征,如声强变化(即语音包络)^[6-11].

脑电信号分析是对脑电图数据进行研究和处理的过程,旨在理解和解释大脑活动的特征和模式.常见的脑电信号分析方法包括时域分析、频域分析、空域分析和时频域分析等.在进行脑电信号分析时,需要综合运用各种方法和技术,结合研究目的,来解读脑电信号的特征和模式.

基于脑电信号的语音感知评估方法通过分析脑电信号的变化,了解不同语音质量对大脑活动的影响,分析语音服务的感知和情绪状态,了解用户的需求和偏好,以优化语音业务的质量和用户体验,进而提升用户的满意度.这种方法为语音感知评估方法的研究提供了一种新的视角和方向.

2 实验方法设计

2.1 实验室配置

本实验采用 64 导的 EEG 放大器(neuracle 博睿康)采集并记录,采样频率为 1 kHz.EEG 电极帽包含 64 个活性电极(neuracle 博睿康),并按照国际 10-20 系统位置分布.接地电极 AFZ 位于电极 AF3 和 AF4 之间,CPZ 电极为参考电极.

实验环境分为屏蔽室和控制室,被试者在专业的屏蔽室完成所有实验步骤.屏蔽室可以屏蔽电磁干扰并严格控制外部噪声、室内温湿度和光线亮度.实验过程中被试者禁止使用手机等其他设备以防止注意力分散,实验由被试者独立完成.测试人员可在控制室通过监控观察屏蔽室内情况,如有特殊情况发生,测试人员可以通过麦克风与被试者进行必要的交流及指引.

2.2 评估语音素材制作

在研究大脑对自然语音加工过程中,理解其神经活动是一个具有挑战性的问题.自然语音不仅包含声学特征,如声音的强度和音调,还承载了丰富的语言信息,如词汇和短语.这使得描述大脑对自然语音的神经活动变得相当复杂.因此,如何准确描述这个过程被认为是认知科学研究领域的一大挑战.

为此,本研究采用自然朗读的方式来呈现语音材料.一名专业的男性播音员在录音室中,对实验语句进行自然朗读,并进行录音.在录制过程中,播音员对实验目的和不同语句之间的差异一无所知,以确保语音的自然性和真实性.该方法能够更真实地模拟日常生活中的听觉场景,为研究大脑在复杂、高生态效度的听觉场景中的听觉注意机制提供更准确的依据.

在自然朗读的语音中,音节的出现并不是按照固定的速率进行的,此外,语音还包含词汇的韵律特征.用于测试的语音长度为 8 s,此长度可以确保实验的稳定性和可靠性,同时也能更好地模拟日常生活中的听觉场景.该语音包括两句日常交流中常用的高频用语,易于理解和交流.每句话之间都有 1~2 s 的间隔,以保证实验材料的合理性和准确性.此外,本语音符合 ITU-T P.863 的相关规定,作为本实验的标准语音材料.

2.3 参与者招募和样本特征描述

本次实验共有 16 位受试者参与,由于测试用语为汉语,因此受试者的母语需为汉语.为了尽量降低实验结果受主观认知的影响,要求所有受试者均未参与过类似的音频实验.受试者的年龄分布在 20~29 岁之间,平均年龄为 23 岁,身体机能、认知都处于成熟阶段.受试者中包括 8 名女性和 8 名男性,均为右利手,听力和视力正常,且无神经系统疾病.每位受试者都成功完成实验,他们的脑电图数据都被完整采集.此外,所有实验程序均已获得本研究所在学校的研究伦理委员会的批准.在实验开始之前,每位受试者都签署了知情同意书,并在实验完成后获得了适当的报酬.

2.4 实验设计和流程

如表 1 所示,级别 1~5 表示在语音通信过程中经过传输后的接收方所听到的语音,不同级别的语音反映了不同的语音质量,其中,级别 5 最接近传输前发送方说出的语音.

实验流程如图 1 所示,在实验过程中,受试者聆听了 120 条标准语音.所有的降级语音以随机顺序呈现,确保不同受试者之间的平衡,避免语音顺序和内容对实验结果的影响.实验要求受试者闭上眼睛,集中注意力听语音,并理解其中的内容.听完每段语音后,受试者需要按表 1 要求对语音进行评分.这一步骤是为了收集受试者对语音清晰度和可理解度的主观感受,从而更全面地评估降级语音的效果.通过这种方式,可以更准确地了解受试者在自然朗读语音中的听觉注意机制,为相关领域的研究提供更有价值的科学依据.

2.5 数据采集和预处理

脑电活动的变化极为迅速,为了准确捕捉这些变化,脑电仪提供了高采样频率,确保脑电信号具有出色的时间分辨率.根据过去的研究,脑电信号主要分布在 5 个不同的频段: alpha 频段((7, 12] Hz)、beta 频段((12, 30] Hz)、delta 频段(不大于 4 Hz)、theta 频段((4, 7] Hz)和 gamma 频段(大于 30 Hz).然而,在记录脑电信号的过程中,存在一些噪声干扰,如低频漂移和工频干扰等.为了确保分析的准确性,本文将主要关注 1~40 Hz 频段的脑电信号.

本实验所有操作均由 Matlab 操作实现.首先改变参考电极为全脑平均,然后使用 FIR (Finite Impulse Respons, 线性相位)滤波器对脑电信号进行带通滤波,通带范围为 1~40 Hz.由实验人员人工进行检查,观察实验信号是否存在肌电、眼电、电极故障等大幅漂移的重大信号干扰现象,一经发现则剔除,避免污染实验数据.经过人工检查清洗之后,采

表 1 语音质量级别
Table 1 Voice quality levels

级别	质量等级	降级量	效果说明
5	优秀	正常	非常好,听得很清楚,无延迟感,无失真感
4	良好	轻微弱化	大多数可以听清楚,延迟小,杂音很少
3	中等	一定程度弱化	有杂音,听不太清楚,基本可以获取语义
2	次等	明显弱化	听不清楚,有杂音,大多数需要努力识别
1	差	无法接受	出现吞字,噪声很大,完全无法理解语义

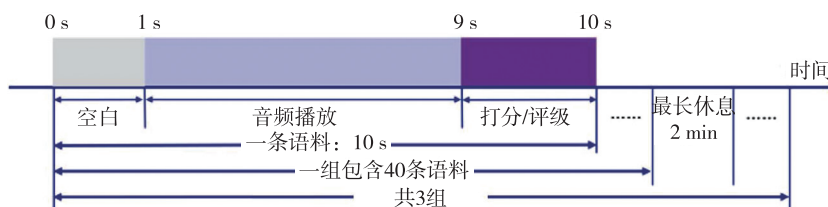


图 1 实验流程
Fig. 1 Experiment flow chart

用 ICA (Independent Component Analysis, 独立成分分析) 去除脑电信号中由于眨眼产生的无意识干扰噪声. 预处理完成后, 截取每个语音开始播放到播放结束后 2 s (共计 10 s) 作为有用信号, 进行下一步分析.

2.6 数据分析方法

相位同步分析方法是将相位分量与振幅分量分离, 以便挖掘脑电信号中隐藏的重要信息. 相锁值是测量相位同步程度的量化值, 表示在一个窄频带内两个信号之间的同步程度, 这种程度是对神经元或神经群进行调节实现的精确锁定^[12-17]. 锁相位算法流程如图 2 所示, 通过希尔伯特变换中的经验模态分解, 逐步分解不同尺度的波动, 并产生几个数据序列, 每个数据序列分布在一个窄频带上, 通过计算瞬时相位进而得到锁相位的值.

1) 经验模态分解

利用 EMD (Empirical Mode Decomposition, 经验模态分解) 对信号进行自适应处理, 是基于其自身尺度特性的一种方法, 目标是实现信号的平稳化. 通过这种方法, 可以将非线性和非平稳的信号分解为有限数量的 IMF (Intrinsic Mode Function, 固有模态函数). IMF 的数量与信号的复杂性相对应, 而每个 IMF 都包含原始信号在不同时间尺度的局部特征信息.

对于一维信号 $x(t)$, 需要识别的是其局部极值点, 包括极大值和局部极小值. 通过样条插值方法对每个极值点进行曲线拟合, 即可得出上包络线 (由局部极大值组成) 和下包络线 (由局部极小值组成). 计算上下包络线的均值, 得到平均包络 $m(t)$, 然后将平均包络与原始信号进行差值运算, 从而得到中间信号 $h(t)$ ^[18]. 如式 (1) 所示:

$$h(t) = x(t) - m(t). \quad (1)$$

假设中间信息 $h(t)$ 满足 IMF 的条件, 将 $h(t)$ 记作一个 IMF, 并计算剩余信号 $r(t)$, 其计算方式如下:

$$r(t) = x(t) - f(t). \quad (2)$$

如果 $h(t)$ 不满足 IMF 的条件, 则假设 $h(t)$ 作为新的 $x(t)$, 并重复上述步骤, 直到 $h(t)$ 满足 IMF 的要求^[19]. IMF 的要求为: 中间信号 $h(t)$ 中不能存在负的局部极大值和正的局部极小值. 使用符号 D 来表示停止条件:

$$D = \frac{\sum_{t=0}^T |h_k(t)|^2}{\sum_{t=0}^T |f_k(t)|^2}. \quad (3)$$

重复上述筛选过程, 得到 k 个 IMF. 当残余分量 $r(t)$ 所包含的极值小于 3 时停止筛选. 式 (4) 为原始

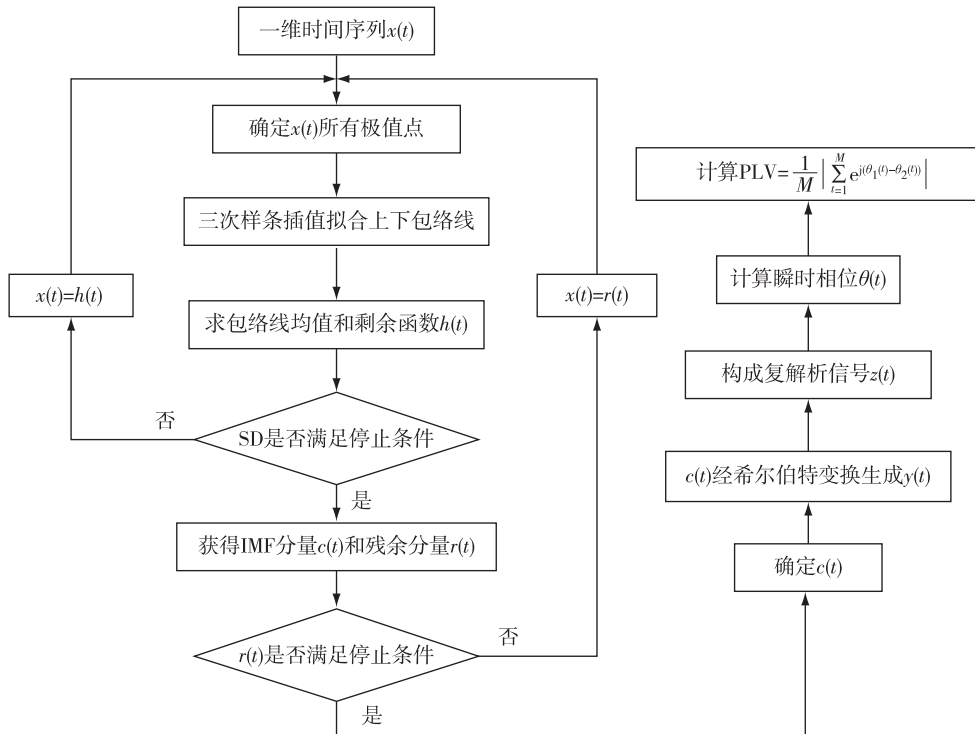


图 2 锁相位流程

Fig. 2 Phase locking flowchart

信号的组成.

$$x(t) = \sum_k f_k(t) + r(t). \quad (4)$$

2) 瞬时相位

基于希尔伯特变换计算瞬时相位^[20]. IMF 的希尔伯特变换记作 $F(t)$:

$$F(t) = \frac{1}{\pi} \zeta \int_{-\infty}^{+\infty} \frac{f(\tau)}{t - \tau} d\tau. \quad (5)$$

其中: ζ 为柯西主值; $f(t)$ 为固有模态函数. $f(t)$ 的解析信号为 $Y(t)$, 如式(6)所示:

$$Y(t) = f(t) + j \cdot F(t). \quad (6)$$

瞬时相位计算公式为

$$\theta_i(t) = \arctan \frac{y_i(t)}{c_i(t)}. \quad (7)$$

3) 相锁值

计算信号间的 PLV (Phase Locking Value, 相锁值):

$$PLV = \frac{1}{M} \left| \sum_{i=1}^M e^{j(\theta_1(t) - \theta_2(t))} \right|. \quad (8)$$

如式(8)所示, 相锁值的取值范围介于 0 和 1 之间, 0 表示没有同步, 1 表示完全同步. 同步的程序取决于两个信号的瞬时相位差, 该差值体现了两个信号相位的波动情况, 0 表示相位差大、波动大, 1 表示相位差小、波动小. 通过对脑电信号的相位分离, 可以发现信号的强弱程度与大脑功能连通性之间的关系^[21].

3 结果分析

3.1 结果分析和讨论

预处理后的脑电信号去除坏段后, 将脑电信号按每段 10 s 进行划分. 在频谱分析时, 先将脑电信号进行平均, 再进行 DFT (Discrete Fourier Transform, 离散傅里叶变换) 获取脑电频谱, 频域分辨率按 1 Hz 计算. 将脑电频谱的能量在脑电电极和受试者之间进行平均.

用于测试的语音波形如图 3 所示, 共包含两句话. 图 4 为频域分析结果(图 4 中, ERSP 表示事件相

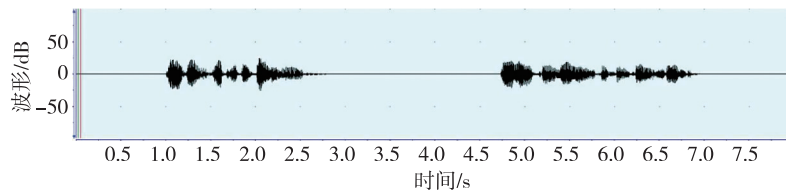


图 3 标准语音波形

Fig. 3 Standard speech waveforms

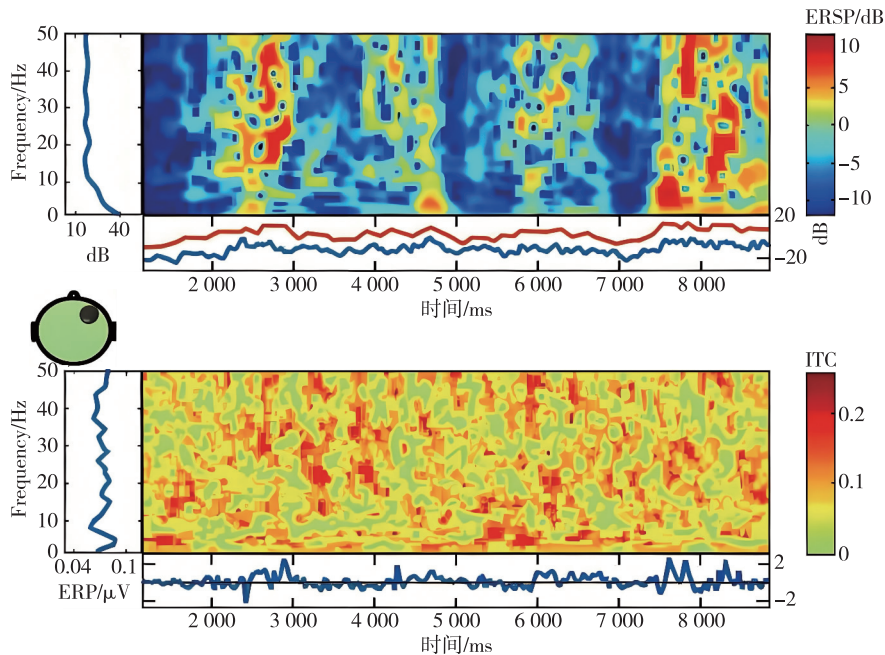


图 4 时频分析

Fig. 4 Time-frequency analysis

关频谱扰动 (Event-Related Spectral Perturbation), ITC(Inter-Trial Coherence, 试验间一致性) 表示事件相关锁相, 蓝色线表示未受语料刺激时的基准线, 红色线表示收听语料刺激后的 ERSP 包络线, ERP 表示事件相关电位 (Event-Related Potential), 单位 μV . 对比图 3 和图 4 可以看出频域图和语音较为吻合. 1 s 开始有语音, 2.5 s 开始出现较强的 beta 和 theta 波, 并产生认知, 在 4.5 s 左右出现分散的 beta 和 theta 波, 并对第一句话产生认知, 第二句话在 6 s 左右开始有较分散的 beta 和 theta 波, 并产生认知, 在 7.5 s 时开始有较强的 beta 和 theta 波, 并产生认知, 由于包括两句话, 因此认知信号更为强烈.

图 5 中, 横坐标表示事件持续时间, 纵坐标表示相关电位的值. 对比图 3 可以看出, 图 5 和语音吻合较好, 1 s 开始出现语音, 在 2.5 s 第一句结束后产生较强负波, 随后第二句话在 4.7 s 开始, 准备接受第二段话刺激, 说话结束后 7.5 s 出现更强负波, 由于认知层面涉及两句话, 因此波幅更大, 且认知集中在前额叶认知区域.

3.2 脑电信号与语音刺激之间的关系

实验中, 受试者被要求听取标准语音刺激, 并同时记录脑电数据, 发现不同人听到相同标准语音时, 有相同的脑电反应趋势.

本文从大脑功能连接的角度去分析脑电信号与语音质量之间的关系. 通常采用互信息方法分析两个信号之间的依赖或相关程度, 计算大脑的功能连

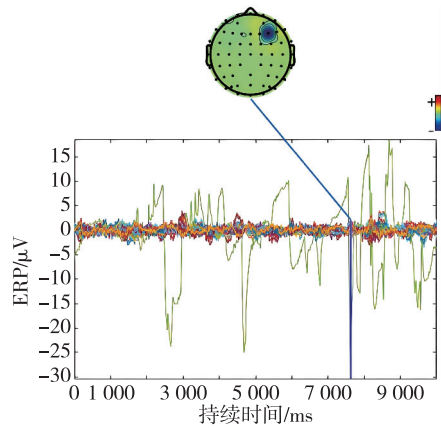


图 5 通道 ERPs

Fig. 5 Channel ERPs

接性. 互信息不仅可以检测信号间的线性相关性, 还可以检测信号间的非线性相关性, 但相关性的准确度受信号的影响较大, 特别是含有噪声和长信号的情况. 为此, 本文选择相锁值的方法, 通过对两个信号进行相位差测量实现功能连接性的分析.

通过经验模态分解, 原始信号得到的固有模态函数分量波形如图 6 所示 (纵坐标为振幅, μV).

计算各个固有模态函数的功率谱, 得到固有模态函数分量的频谱, 如图 7 所示 (纵坐标为功率谱, dB).

由于经验模态分解出的固有模态函数分量有多个, 故需要选定某个固有模态函数分量进行后续的分析. 通常选取前几阶的 IMF 分量再平均进行分析,

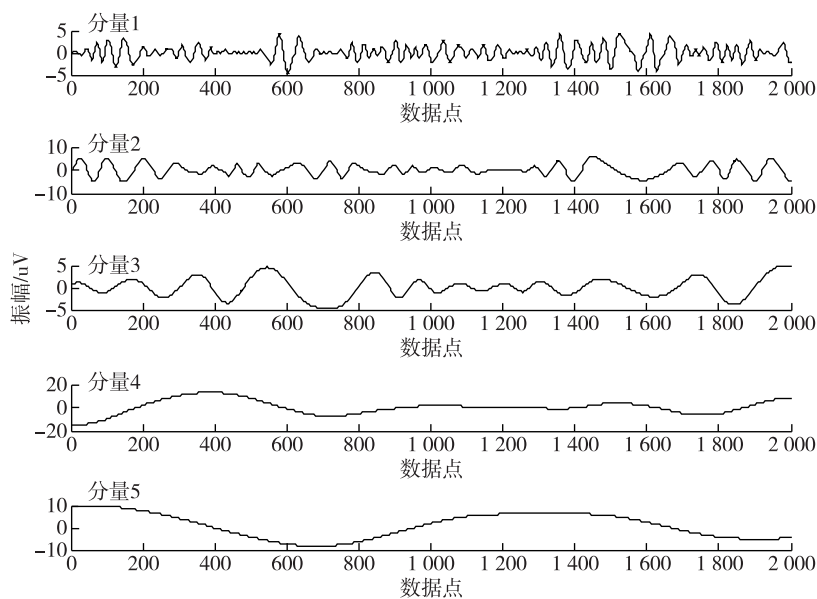


图 6 固有模态函数分量

Fig. 6 IMF components

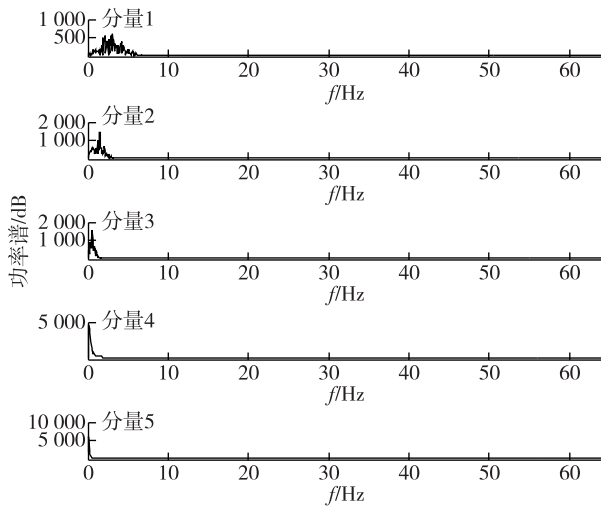


图7 固有模态函数分量频谱

Fig. 7 Spectrogram of IMF components

这种方法不具有自适应性.为了解决这个问题,本文提出一种改进锁相位的方法(图8).以距离作为衡量值,计算每个固有模态函数分量与原始信号间的距离,体现分解信号与原始信号的相似性.本文引入DTW(Dynamic Time Warping,动态时序相似度)度量算法作为标准,选取固有模态函数分量.

DTW 是一种距离计算方法,用来检测两个时序的相似程度.距离越小,相似性越大,故以最短距离的固有模态函数分量作为新信号,再计算各个通道之间的相锁值,实现对皮层脑电的功能连接性分析.

为了更好地体现各个电极之间的同步性,选取不同脑区具有代表性的电极:O1、O2、P3、P4、Pz、P7、P8、C3、C4、Cz、T7、T8、F3、F4、Fz、F7、F8、FP1、FP2.这19个通道分别对应着枕区、顶区、中央区、颞区、



图8 改进锁相位流程

Fig. 8 Block diagram of improved phase locking flow

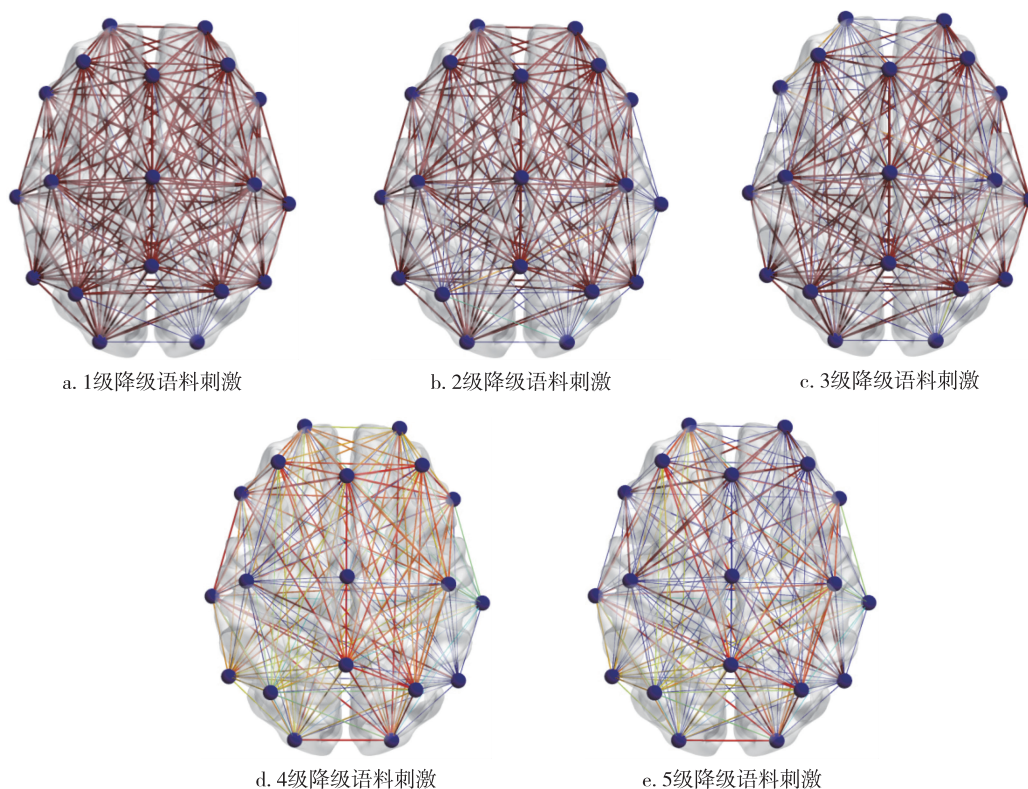


图9 基于相锁值的大脑功能连接性分析

Fig. 9 Brain functional connectivity analysis based on phase locking values

额区和前额区.图 9a—e 分别对应 1、2、3、4、5 级降级语料刺激下的大脑功能连接示意图.图 9a 的红色线段最多,说明 1 级降级语料的刺激下大脑功能连接性增强,表示需要较为集中地去思考听到的内容;图 9b、9c 的红色线段有所减少,且图 9c 比图 9b 更少,说明 2、3 级降级语料的刺激下大脑功能连接性降低,能够大约听清楚语料的内容;图 9d 的橙黄色线段增加,说明 4 级降级语料的刺激下大脑功能连接性再次降低,能够较为清楚地听到语料内容;图 9e 的蓝色线段增至最多,说明 5 级降级语料的刺激下大脑功能连接性最弱,表明可以清晰地听到语料内容,无需认真思考语料所表达的内容.不同降级语料下的脑电相锁值统计结果表明,脑电信号的功能连接程度存在显著差异($P < 0.05$).

将 5 种级别的语音所生产的大脑功能连接图作为输入,经过深度学习网络验证是否能有效区分不同的级别.深度学习网络的模型为四层卷积,两层卷积之间为最大池化层,两个全连接层中间为一个 Dropout 层,最终的分类正确率为 99.62%.

实验结果表明,在听到相同标准语音时,被试者们的脑电信号呈现一致的反应趋势.同时,该发现也为一种与语音刺激内容相关的神经认知机制提供了依据.

4 结论和展望

本研究的应用背景为无线多媒体通信,相关技术的高速发展导致使用传统的客观质量评价难以衡量用户的体验质量.如果采用基于用户主观反馈的评估方法,则过度依赖用户,且容易受到用户的高层认知偏差影响.而使用脑电信号分析用户对于音频的感知过程,可以有效地消除上层认知偏差.本文设计的一种基于自然刺激脑电实验,采用符合 ITU-T P.863 的标准语音作为刺激材料,组织 16 位受试者参与测试,通过实验得出所有受试者在听取相同的标准语音的脑电反应上展现出一致性的趋势.并且,脑电信号与语音质量之间具有功能连接性,脑电信号正确区分语音质量的正确率达到 99.62%.

研究结果表明,使用脑电信号对音频质量评估具有可实现性、可解释性,证明了脑电信号更能反映用户的真实体验质量,为之后使用脑电信号评估用户体验质量提供了依据.在今后的研究中,可以利用源定位等技术手段,对脑电信号产生的位置进行标记,以提取更准确的脑电信号,减少误差.此外,除了

宋奇蔚,等.基于语料语义的脑电反应一致性研究.
SONG Qiwei, et al. A study of EEG response coherence based on corpus semantics.

脑电信号,还可以将磁共振、脑磁图、心电图、眼动仪等多种手段结合,提出多模态的主观质量评估方法,为主观体验量化进行科学性描述,最终达到准确评估业务质量的目的.

参考文献

References

- [1] Bar O, Shallom I D. Speech quality assessment using EEG signals [C] // 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE). November 16–18, 2016, Eilat, Israel. IEEE, 2016: 1–5
- [2] Uhrig S, Möller S, Voigt-Antons J N. Dissociating perceptual quality dimensions of transmitted speech using electroencephalography [C] // 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX). May 29–June 1, 2018, Cagliari, Italy. IEEE, 2018: 1–3
- [3] Recommendation ITU-T P.863. 1 (06/2019): methods for objective and subjective assessment of speech and video quality [EB/OL]. [2024-07-01]. <https://www.itu.int/rec/T-REC-P.863/>
- [4] Keitel A, Gross J, Kayser C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features [J]. PLoS Biology, 2018, 16 (3): e2004473
- [5] Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex [J]. Neuron, 2007, 54 (6): 1001–1010
- [6] Di Liberto G M, O'Sullivan J A, Lalor E C. Low-frequency cortical entrainment to speech reflects phoneme-level processing [J]. Current Biology, 2015, 25 (19): 2457–2465
- [7] Ding N, Simon J Z. Emergence of neural encoding of auditory objects while listening to competing speakers [J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109 (29): 11854–11859
- [8] Destoky F, Philippe M, Bertels J, et al. Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope [J]. NeuroImage, 2019, 184: 201–213
- [9] Nourski K V, Reale R A, Oya H, et al. Temporal envelope of time-compressed speech represented in the human auditory cortex [J]. The Journal of Neuroscience, 2009, 29 (49): 15564–15574
- [10] Peelle J E, Gross J, Davis M H. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension [J]. Cerebral Cortex, 2013, 23 (6): 1378–1387
- [11] Koskinen M, Seppä M. Uncovering cortical MEG responses to listened audiobook stories [J]. NeuroImage, 2014, 100: 263–270
- [12] Lachaux J P, Rodriguez E, Martinerie J, et al. Measuring phase synchrony in brain signals [J]. Human Brain Mapping, 1999, 8 (4): 194–208

- [13] Allefeld C, Kurths J. Multivariate phase synchronization analysis of EEG data [J]. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2003, E86-A(9): 2218-2221
- [14] Allefeld C, Kurths J. An approach to multivariate phase synchronization analysis and its application to event-related potentials [J]. *International Journal of Bifurcation and Chaos*, 2004, 14(2): 417-426
- [15] Rosenblum M G, Pikovsky A S, Kurths J. Phase synchronization of chaotic oscillators [J]. *Physical Review Letters*, 1996, 76(11): 1804-1807
- [16] Li D, Li X L, Cui D, et al. Phase synchronization with harmonic wavelet transform with application to neuronal populations [J]. *Neurocomputing*, 2011, 74(17): 3389-3403
- [17] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. *Proceedings of the Royal Society of London Series A*, 1998, 454(1971): 903-998
- [18] 李昕, 范梦颀, 孙小棋, 等. 基于改进锁相位算法的脑区功能同步性研究 [J]. *生物医学工程学杂志*, 2018, 35(3): 350-357
LI Xin, FAN Mengdi, SUN Xiaoqi, et al. Research on the correlation of brain function based on improved phase locking value [J]. *Journal of Biomedical Engineering*, 2018, 35(3): 350-357
- [19] 陈东毅, 陈建国, 李玉榕. 结合经验模态分解能量矩占比和方差贡献率法的步态信号特征向量提取 [J]. *物联网技术*, 2019, 9(2): 35-39, 42
- [20] Zoefel B, Huster R J, Herrmann C S. Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance [J]. *NeuroImage*, 2011, 54(2): 1427-1431
- [21] Aviyente S, Bernat E M, Evans W S, et al. A phase synchrony measure for quantifying dynamic functional integration in the brain [J]. *Human Brain Mapping*, 2011, 32(1): 80-93

A study of EEG response coherence based on corpus semantics

SONG Qiwei¹ GE Ning¹ XIANG Tao¹

¹ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Abstract This study examined the response tendency of the human brain to a standardized corpus of speech by measuring changes in speech-related electroencephalographic (EEG) signals. Sixteen participants listened to 120 standardized speech items, each lasting 8 seconds, with intervals of 1 to 2 seconds between them and played in a random order. During the listening process, the EEG signals were extracted from the participants, and the signals within the frequency band of 1-40 Hz were preprocessed and analyzed in comparison with the speech signals. The results showed that participants exhibited similar EEG response trends when exposed to the same standardized speech. Furthermore, phase difference analysis between EEG and speech signals was conducted using the phase locking value method, which demonstrated the functional connectivity between EEG signals and speech quality. Notably, the EEG signals achieved a 99.62% accuracy in distinguishing speech quality.

Key words EEG; speech quality; perceptual assessment; natural stimulation