

基于汉语音位发音想象的脑机接口研究*

杨晓芳^{1,2,**}, 江铭虎^{1,2}

(1. 计算语言学实验室, 清华大学人文学院, 北京, 100084, 中国;

2. 心理学与认知科学研究中心, 清华大学人文学院, 北京, 100084, 中国)

摘要: 本文提出了一个基于汉语音位发音想象的脑机接口系统框架, 使得受试者使用脑机接口系统时能更加自然和流畅。三名受试者参与了本实验研究, 实验过程中受试者被要求想象四个汉语元音和四个辅音音位的发音部位及语音发音, 以及一个不作想象任务的控制条件, 同时记录其脑电数据。在数据处理阶段, 本文对采集到的头皮脑电数据进行了频域、时域、空域分析, 以提取出音位发音想象效应最优化的特征向量用于提高每两个条件间的配对分类效果。实验结果表明, 音位发音想象效应的最优脑电频段为 2~10 Hz, 时段为刺激呈现后 300~500 ms, 头皮空间分布主要集中在感觉运动皮层区域。音位发音想象任务和控制条件相比具有较高的分类正确率, 最高可达 83%, 为基于音位发音想象的汉语脑机接口系统研究提供了理论基础。此外, 刺激材料间的 Jaccard 距离和分类正确率的高度相关性表明, 音位发音想象任务可被视为复杂的发音器官运动想象任务, 并且可由人脑感觉运动皮层区域的脑电信号来解码预测。

关键词: 脑机接口; 运动想象; 音位发音

中图分类号: TP391

文献标识码: A

Study on Brain-Computer Interface Based on Chinese Phoneme Imagery

Xiaofang Yang^{1,2}, Minghu Jiang^{1,2}

(1. Lab of Computational Linguistics, Tsinghua University, Beijing, 100084, China;

2. Center for Psychology and Cognitive Science, Tsinghua University, Beijing, 100084, China)

Abstract: This paper intends to construct the framework of a voluntary speech neural prosthesis based on phoneme imagery EEG signals to make brain-computer interface (BCI) speech production more natural and fluent. EEG signals were recorded in three healthy subjects while imagining both the vocalization and places of articulation of four vowels and four consonants in Mandarin Chinese as well as a no imagination state as control. To process the EEG data, this study performed spectral, temporal, and spatial analyses to extract the optimal phoneme imagery features for pairwise classification by SVM between every two tasks. The results revealed that the phoneme imagery effect was demonstrated in the frequency range of 2~10 Hz, the time interval of 300~500 ms after the stimuli onset, and the spatial patterns with strong activities mainly covering the sensorimotor cortical region. Besides, this study also found a high correlation of the pairwise classification accuracies with the Jaccard distances between the experimental stimuli based on the binary descriptions of articulation control. This experiment confirms the hypothesis that phoneme imagery can be characterized as a complex motor imagery task and that, with the highest classification accuracy between speech imagery and non-imagery tasks reaching up to 83% averaged across subjects, scalp-level speech motor imagery signals probably possess an unfulfilled potential to control a neural utterance synthesizer for communication BCIs.

Key words: Brain-computer Interface (BCI); Motor Imagery; Phoneme Articulation

* **基金项目:** 国家自然科学基金项目 (61171114); 清华大学自主科研项目 (20111081023 & 20111081010)

** **通讯作者,** Email: xfyang.thu@gmail.com

1 引言

脑机接口 (Brain-Computer Interface, BCI) 技术是近年来生物医学工程领域的研究热点之一, 形成于 20 世纪 70 年代^[1]。1999 年, 第一届 BCI 国际会议给出了 BCI 的定义, 即“脑机接口是一种不依赖于正常的由外周神经和肌肉组成的输出通路的通讯系统”^[2]。传统的人机交互方式通过声音、按钮等形式实现, 然而部分病人因各种原因不具备言语表达或肢体操作能力, 故此类交互方式无法实现。幸运的是 BCI 技术为其提供了一种全新的通信和控制方式, 无需通过语言或肢体动作, 而是直接通过脑电波来表达其想法或操纵设备, 可帮助罹患中风或肌萎缩侧索硬化症 (Amyotrophic Lateral Sclerosis, ALS) 等运动神经元疾病的患者部分或完全恢复与外界沟通交流的能力^[1], 也可以为健康人提供全新的沟通及娱乐方式^[3]。

随着微导线 (microwire) 和微电极阵列 (microelectrode array) 的研发与应用, 研究人员能够通过皮层脑电图 (Electrocorticogram, ECoG), 局域场电位 (Local Field Potentials, LFPs), 以及单单元动作电位 (Single Unit Activity, SUA) 等方式记录到受试者的颅内电生理信号, 继而通过信号处理和模式识别等手段解码高时空分辨率颅内脑电信号, 从而解读脑机接口系统使用者的意图并由此控制外部设备按其意愿工作, 例如移动鼠标至指定位置并选中打开某程序, 通过脑电信号打字向电脑屏幕输入想说的内容, 甚至是控制轮椅向指定方向移动或是机械假肢按指令行动等^[4,5,6]。尽管颅内脑电信号具有很高的时间和空间分辨率, 但其通过开颅手术放置电极而造成对人体损伤的有创性使之很难得以推广普及。因而, 为使脑机接口系统面向更广大的用户群体, 研究人员对于无创的 BCI 系统进行了大量研究。通过脑电图 (Electroencephalogram, EEG) 和脑磁图 (Magnetoencephalogram, MEG) 等手段测量脑部活动信号从而解读使用者的意图^[7,8,9,10]。在无创脑机接口系统研究中, 研究最广泛的当属基于 P300 成分的脑机接口系统, 通过记录使用者的头皮脑电信号来测量其脑部活动, 并探测 P300 成分诱发信号以实现人脑同外部设备的直接交互。P300 脑电成分具有被诱发或未被诱发两种状态, 其诱发条件为主动注意某一刺激, 低层次感知特性^[11,12], 或是由低频率呈现的刺激所诱发^[13,14]。由此, P300 成分可令使用者在未接受特定训练的情况下通过调整自身注意力, 使其集中在某一特定刺激上而达到运用头皮脑电信号与外部设备进行沟通的目的, 从而控制脑机接口系统按其意愿行事。

然而, 基于 P300 成分的脑机接口系统依赖外部设备提供刺激输入, 并要求使用者长时间集中注意力, 这一要求对某些使用者尤其是病人较难实现。研究人员继而转向基于感觉运动皮层节律 (Sensorimotor Rhythm, SMR) 的 BCI 系统, 通过判别不同的运动想象 (Motor Imagery, MI) 模式来实现人脑和外部设备的交互^[15,16,17]。大脑在产生动作意识或受到外界刺激时, 神经细胞将产生十分微弱的电活动, 大量神经细胞的电活动传到头皮表层形成脑电波, 此 EEG 信号体现出某种节律和空间分布的特征, 并且可以通过一定的方法加以检测, 再通过信号处理与模式识别等手段从中辨析出人的意图信号, 进而将其转换为控制命令, 以实现对外部设备的控制和与外界的交流^[18]。因此, 为使 BCI 用户能通过更加自然流畅的人际沟通方式主动控制脑机接口系统, 而不仅仅是被动接受外部刺激输入, 本文假设发音想象任务可被视为复杂的发音器官运动想象任务, 并提出了一种基于音位发音想象的自主语言神经假体控制方案, 通过脑电实验验证其可行性。实验过程中受试者被要求想象汉语元音和辅音音位的发音部位及语音发音, 同时记录其脑电数据。在数据处理阶段, 本实验对记录到的脑电信号进行了频域、时域、空域分析以提取出音位发音想象效应最优化的特征向量来提高脑电数据的分类效果。

本文剩余部分按照如下方式组织: 第 2 节详细介绍了汉语音位发音想象脑电实验的实验范式、实验材料、数据采集、数据处理、时频分析、共同空间模式算法以及脑电数据配对分类; 第 3 节描述了实验结果, 包括频域分析结果、时域分析结果、空域分析结果和配对分类结果; 最后, 第 4 节是全文总结与讨论。

2 脑电实验

2.1 实验范式

三位来自清华大学的本科生参与了此次实验（两位男性，平均年龄 = 19.3 岁，标准差 = 2 岁）。所有受试者均来自中国大陆，并且都能说标准流利的普通话。所有受试者均有正常或矫正至正常的视力，且根据爱丁堡利手测验均是右利手^[19]。没有受试者报告有任何神经或心理方面的疾病，且之前均没有任何使用脑机接口系统的经验。整个实验过程完全符合赫尔辛基宣言，且所有受试者均在实验前签署了知情同意书。

受试者头戴 EEG 电极帽非常放松地坐在光线较暗的电磁屏蔽室里的扶手椅上，距离刷新率为 60 赫兹的 CRT 显示屏大约 60 厘米。在实验开始之前，受试者均接受了汉语发音训练确保其发音部位和方法正确。实验过程中受试者被要求想象屏幕中央呈现的汉语音位的发音部位及语音发音。在实验开始时 ($t = 0$ s)，电脑灰色屏幕中央出现一个固定的十字提示本次实验即将开始。一至两秒过后，屏幕中央随机出现一个刺激符号并停留两秒钟，受试者根据提示想象对应的音位发音直至该刺激符号消失才停止。之后电脑屏幕再次变灰并持续三秒钟供受试者休息，如此便完成了一次实验。具体实验流程如图 1(a)所示。选择灰色屏幕作为背景是为了降低视觉后效。每个刺激符号进行 50 次实验，每位受试者共进行 450 次实验。

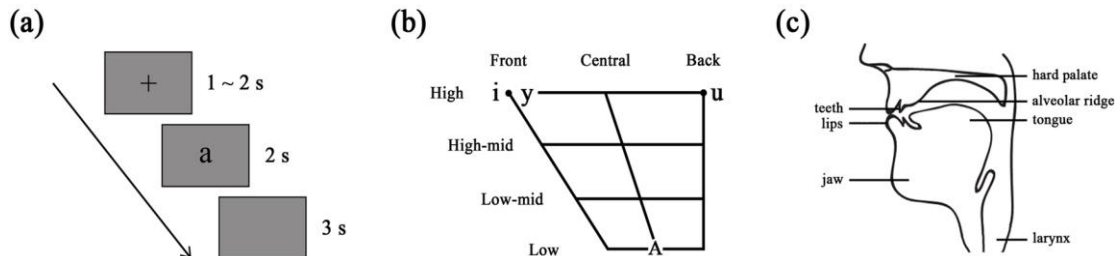


图 1. (a) 单个试次实验流程图；(b) 普通话元音舌位图；(c) 发音部位矢状位解剖图

2.2 实验材料

本实验过程中共有九种刺激材料呈现在屏幕中央，字体均为 SimSun-ExtB，大小为 30。其中四种是元音音位，分别代表普通话四呼的一类，即开口呼 (/a/)、齐齿呼 (/i/)、合口呼 (/u/) 和撮口呼 (/y/)。表 1 描述了这四个元音音位的发音部位及发音方法。图 1(b)根据国际音标表描绘了普通话元音舌位图中这四个元音音位最具代表性的四个音素，其中水平线表明舌位前后，垂直线表明舌位高低，位于圆点左侧的是非圆唇音，右侧的是圆唇音。需要强调的一点是，音位作为音系学的最基本单位通常两边用斜线表示，音素作为语音学的最基本单位通常两边用方括号表示。在某一特定语言系统内部，音位是由一组彼此差别没有区别词的语音形式作用而音感上又相似的音素概括而成的音类。抽象的音位以具体的音素与音素之间的关系作为基础^[20]。例如，英语中 *kin* 和 *skin* 两个词中的 *k* 分别对应国际音标 [k'] 和 [k]，但这两个音素由同一个音位 /k/ 表示。同样地，在汉语中图 1(b)里的音素 [A] 即属于表 1 中的音位 /a/。由于细微的舌位差别会导致不同的音素发音，本文着重研究音位的发音想象，允许不同试次和受试者间细微的发音差别。

另四种刺激材料是辅音音位，表 1 描述了这四个辅音音位的发音部位及发音方法，各发音部位的矢状位如图 1(c)所示。本文挑选的这四个辅音音位不仅发音部位各不相同，而且发音时长都相对较长。鼻辅音（如 /m/, /n/, /ŋ/）发音时软腭下垂使气流从鼻腔泻出，而摩擦音（如 /f/）发音时气流被迫从发音器官形成的狭窄缝隙中流出。相反，爆破音（如 /b/, /d/, /g/）发音时则是先由发音器官在口腔中形成阻碍，然后气流冲破阻碍而发出声音，因而发音时长较短，不符合本文实验设计要求，因受试者被要求保持音位发音想象两秒钟直至刺激材料在屏幕中央消失。

最后一种刺激材料我们选用非语音符号“#”作为控制条件，使得各种刺激材料都以一

致的字体和大小呈现以保持基本一致的视觉刺激效果。在前人研究中，Dassalla 报告了一个所谓的发音想象语言相关 ERP 成分，其可信度尚存争议，因其实验中仅使用了白屏作为控制条件，更多讨论详见第 4 节^[21]。

表 1. 元辅音音位的发音部位及发音方法二值描述

Articulator	Status	/a/	/i/	/u/	/y/	/m/	/f/	/n/	/ŋ/	#
lips	unrounded	1	1	0	0	1	1	0	0	0
	rounded	0	0	1	1	0	0	0	0	0
teeth		0	0	0	0	0	1	0	0	0
tongue	high	0	1	1	1	0	0	1	1	0
	low	1	0	0	0	0	0	0	0	0
	front	0	1	0	1	0	0	1	0	0
	central	1	0	0	0	0	0	0	0	0
	back	0	0	1	0	0	0	0	1	0
jaw		1	0	0	0	0	0	0	1	0
alveolar ridge		0	0	0	0	0	0	1	0	0
hard palate		0	0	0	0	0	0	0	1	0
larynx		1	1	1	1	1	0	1	1	0

2. 3 数据采集

32 个银/氯化银电极根据国际 10-20 系统标准放置在 NeuroScan 电极帽上用于记录 EEG 实验数据（采样率 = 500 Hz，带通滤波 = .01~100 Hz，陷波滤波 = 50 Hz）。两个散电极被分别放置于左眼上下两侧用于测量垂直眼电（VEOG），水平眼电（HEOG）则由放置于双眼两侧的另两个散电极测量。还有两个散电极被放置于左右乳突，脑电信号采集时以左侧乳突作为参考电极。所有电极的电阻均被调至 5KΩ 以下。EEG 数据由运行于 Windows XP 平台上的 Scan2.0（NeuroScan Inc.）软件进行记录。

2. 4 数据处理

所有受试者的 EEG 数据均由 MATLAB(7.14.0, MathWorks, Inc., Natick, MA)和 EEGLAB 处理^[22]。预处理阶段，首先将采集到的 EEG 信号以左右乳突均值作重参考并作带通滤波（1~30 Hz）以去除低频基线漂移或电极噪声。之后，各个试次的的数据被截断为[-200 800] ms 的时间段并对各通道用刺激呈现前 200 ms 的数据作基线矫正。若某一段数据在任一通道超出 ±100 μV 的范围则被剔除以除去由眨眼、水平眼动、电极噪声等引起的伪迹，提高数据质量便于进行后续分析处理。

经过预处理的脑电数据，我们首先对其进行频域分析，将音位发音想象与控制条件作对比，并通过分析事件相关频谱扰动（Event-Related Spectral Perturbation, ERSP）的差异得出音位发音想象效应统计显著的频段。之后，我们对脑电数据用最优频段作带通滤波并对其进行配对 T 检验以得出音位发音想象效应的最优时段。在提取出最优时段特征后，我们运用 CSP（Common Spatial Patterns）算法设计空间滤波器进一步在空域上优化数据以提高每两个条件之间用线性核函数支持向量机（Support Vector Machine, SVM）进行两两配对分类的效果。完整的数据分析和分类流程如图 2 所示。此外，我们基于表 1 中发音控制的二值描述计算出了各刺激材料间的 Jaccard 距离并通过层次聚类将其可视化。我们又进一步对配对分类的正确率和相应的 Jaccard 距离作了相关性分析。

2. 5 时频分析

在抽取音位发音想象效应的特征时，我们将音位发音想象任务和控制条件作对比以得出

最优化的时频脑电特征。我们利用 EEGLAB 提供的快速傅里叶变换 (Fast Fourier Transform, FFT) 进行脑电数据的频域特征分析,并用二维(潜伏期 × 频率)事件相关频谱扰动 (ERSP) 来表示相对于基线的脑电频谱能量平均变化^[23]。我们选用 256 ms 的滑动窗应用了 200 次,输出的潜伏期范围是-72~672 ms,频率范围是 1~30 Hz。我们进一步用 EEGLAB 提供的拔靴统计法 (Bootstrap Statistical Method) 来检验差异 ERSP 的统计显著性 ($P < 0.025$)。

脑电数据的时域特征分析则是通过点对点配对样本 T 检验将音位发音想象任务和控制条件作比较。我们对运动想象脑机接口研究中常用的感觉运动皮层区的 9 个电极 (FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4) 进行分析,当且仅当每个电极有至少 10 个连续采样点统计显著 ($P < 0.05$) 时,该时间段才被认为效应显著。在呈现时域分析结果时,我们进一步将这 9 个电极在点对点配对样本 T 检验中得出的 ERP 时段平均波幅作配对 T 检验以验证该成分的统计显著性。

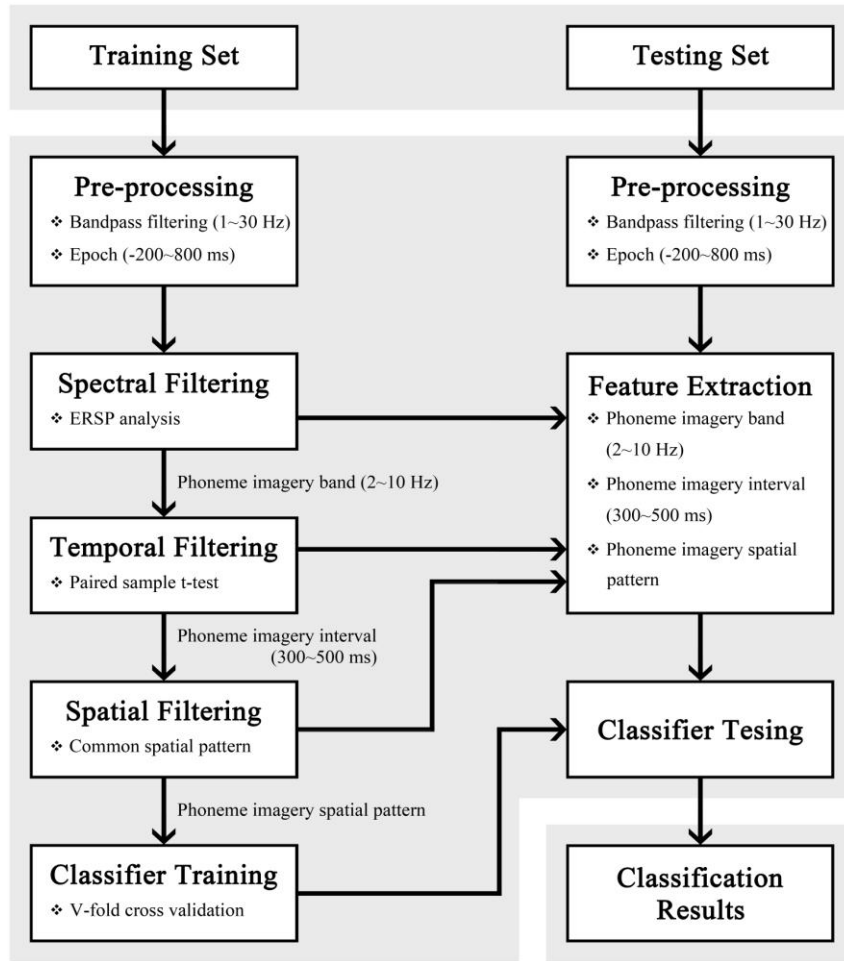


图 2. 脑电数据处理流程图

2. 6 共同空间模式算法

共同空间模式 (Common Spatial Patterns, CSP) 算法被广泛用于提高 EEG 数据的信噪比,通过设计空间滤波器来提取不同实验条件间方差最大的 EEG 数据成分。CSP 算法最初是为两分类脑机接口系统 (Brain-Computer Interfaces, BCI) 设计的^[24],后被扩展至多分类模式^[25]并进一步改进以提供满意的实验结果^[26]。本实验主要针对两两配对分类任务,故本节简要描述两分类 CSP 算法。 $E_g^n (n \in [1, N], g \in [1, G])$ 表示一个试次的原始 EEG 数据矩阵,其中行表示导联,列表示采样数据, n 是 epoch 段落标签,而 g 则是 group 组别标签。各组的平

均空间协方差矩阵 \bar{C}_g 可通过下式得到

$$\bar{C}_g = \frac{1}{N} \sum_{n=1}^N \frac{E_g^n (E_g^n)^T}{\text{trace}(E_g^n (E_g^n)^T)}$$

其中 N 是 g 组内的试次数, $\text{trace}(x)$ 是 x 矩阵对角元素的和。之后, 我们将各组所得的矩阵相加以得到复合空间协方差矩阵

$$C_c = \sum_{g=1}^G \bar{C}_g$$

该矩阵可被分解为 $C_c = U_c D_c U_c^T$, 其中 U_c 是特征向量矩阵, D_c 是特征值对角矩阵。在本节中, 假设特征值均按降序排列。随后, 我们通过下式进行白化变换

$$P = \sqrt{D_c^{-1}} U_c^T$$

以均衡特征空间的方差, 并且将原先的平均协方差矩阵按下式转换

$$S_g = P \bar{C}_g P^T.$$

对于二分类 CSP 算法 (即 $G = 2$), S_g 可以对角化为 $S_g = U D_g U^T$ 使得 $\sum_{g=1}^G D_g = I$,

其中 S_g 拥有相同的特征向量且特征值和为 1, 使得对一类数据特征值最大的特征向量对另一类数据特征值最小, 反之亦然。因此, 我们选取 U 中最前 L 个及最后 L 个特征向量构成空间滤波器。然后, 我们可按下式得到映射矩阵

$$W_g = U_g^T P$$

其中 U_g 由每个类别的 L 个特征向量组成, 而 W^{-1} 的列则是共同空间模式, 即各条件下脑电源分布向量。最后, 我们将各组别每个试次的 EEG 原始数据按下式进行分解

$$Z_g^n = W_g E_g^n$$

以得到新的时间序列, 其方差能最大化地区分各种条件。

2.7 配对分类

本文通过训练支持向量机 (Support Vector Machine, SVM) 来实现每两个条件之间的配对分类, 因为 SVM 具有较强的泛化能力并且对过拟合和维度灾难 (curse-of-dimensionality) 问题有较强的适应能力^[27]。SVM 处理二分问题的基本原理是寻找一个最优化的分类超平面, 使得两类数据间的分类间隔最大化^[28]。而对于非线性 SVM 分类器, 则采用核函数将数据点投射到另一高维空间中, 使其在核空间内线性可分^[29]。

本文对经过预处理的脑电数据用音位发音想象效应的频段和时段提取出最优化的时频特征向量, 并将其降采样至 50 Hz。之后, 从每类刺激材料的 50 段数据中随机挑选 40 个样本组成训练数据集, 剩余样本组成测试数据集, 并用 CSP 算法计算两类训练数据的共同空间模式。EEG 训练和测试数据采用 8 个最重要的空间滤波器 (每一类 4 个) 进行空域分解, 然后再用 LIBSVM^[30]提供的 SVM 线性核函数分类器实现配对分类。分类器首先用训练数据集的特征向量进行训练, 之后再用测试数据集验证其分类效果, 分类器参数则选用网格搜索

和交叉验证中正确率最高的参数^[31]。本文对每位受试者的九种任务进行两两配对（共 36 对组合），每次随机选取训练集和测试集进行分类并重复这一过程 20 次，再将受试间分类平均正确率和每一对刺激材料间的 Jaccard 距离作相关性分析。

3 实验结果

3.1 频域分析结果

本节展示了受试 3 的脑电数据频域分析结果，其余受试者的结果与其相似。受试 3 发音控制皮层区域 C3 和 C4 电极的代表性元音音位 /i/、辅音音位 /f/ 的发音想象、控制条件、音位发音想象效应（/i/ 减去控制条件及 /f/ 减去控制条件）的频谱分析结果和 ERSP 图如图 3 所示，其中音位想象效应的频谱能量由 EEGLAB 提供的拔靴统计法来检验其显著性（ $P < 0.025$ ）。频域分析结果显示，音位想象效应相对于控制条件诱发了双侧感觉运动皮层区 2~10 Hz 频段内的频谱能量降低，这一频段被用来对脑电数据作进一步的频域滤波以提取最优化的频域脑电特征。

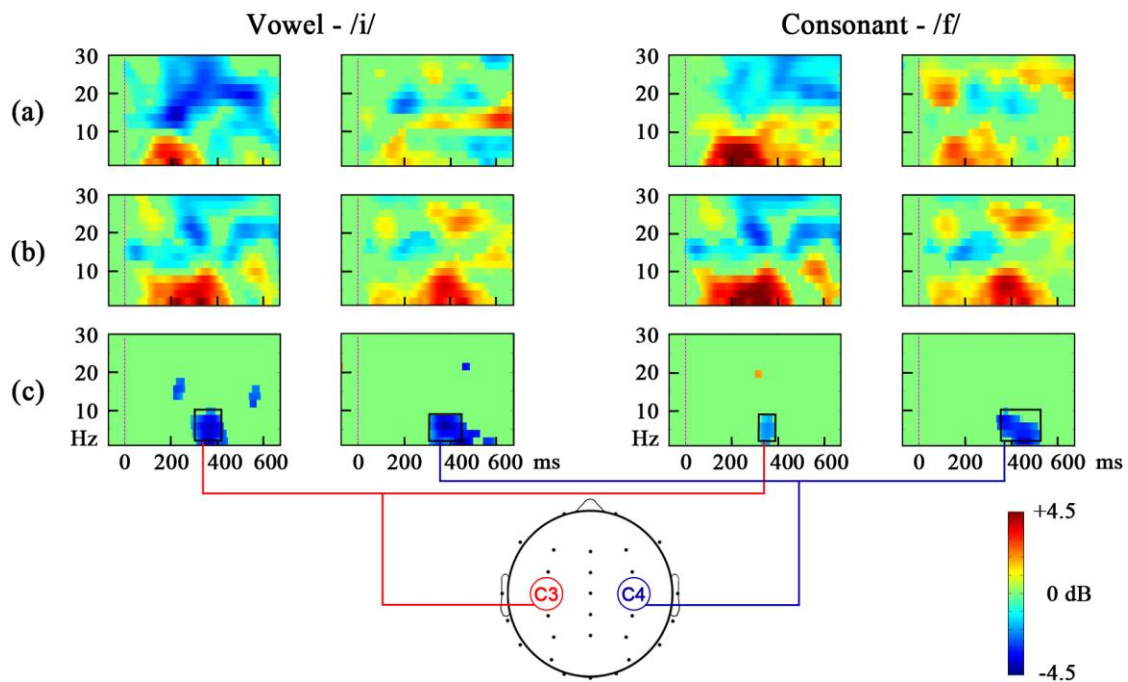


图 3. (a) 发音想象；(b) 控制条件；(c) 音位发音想象效应频谱分析结果

3.2 时域分析结果

时域分析结果得出了 3 个主要时间段，对应不同音位发音想象和控制条件的 ERP 波幅差异。受试间感觉运动皮层区域电极（FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4）平均的音位发音想象及控制条件 ERP 波形如图 4 所示，其中阴影部分表示音位发音想象效应统计显著（ $P < 0.05$ ）的时间段。在早期（98~142 ms）和晚期（446~530 ms）阶段，音位发音想象任务较控制条件相比分别诱发了一个更负向的 ERP 成分，而在中期（308~396 ms）阶段则是诱发了一个更正向的 ERP 成分。音位发音想象任务和控制条件早期的差异很大程度上是由于视觉刺激的物理属性（如空间频率）不同而引起的，因而本文将音位发音想象效应的时间窗缩短至 300~500 ms，涵盖了绝大部分音位刺激材料诱发的中期和晚期脑电响应时段，并用这一 200 ms 的时间窗对脑电数据作进一步的时域滤波以提取最优化的时域脑电特征。

3.3 空域分析结果

本文采用共同空间模式算法计算出 80 个训练样本数据（每类 40 个训练样本）的最优化空间滤波器，并挑选其中 8 个最重要的空间滤波器（每类 4 个空间滤波器）来生成空域最优化的特征向量用于后续分类任务。图 5 展示了在一次 CSP 迭代计算中，受试 3 想象音位 /u/

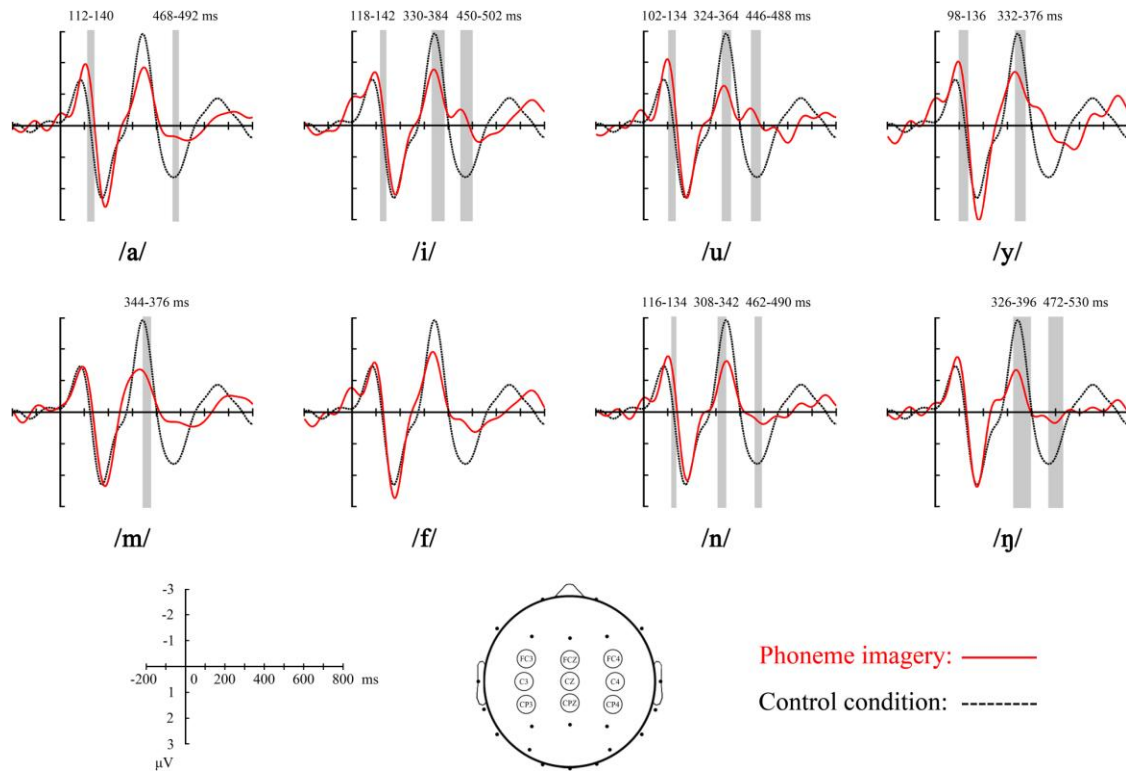


图 4. 感觉运动皮层区域电极平均的音位发音想象及控制条件时域分析结果

的发音和控制条件的 8 个最重要的共同空间模式。第一行包括了音位/u/发音想象任务重要性第一至第四的空间模式 (CSP1~4)，而第二行则包括了控制条件重要性第一至第四的空间模式 (CSP1~4)。音位/u/发音想象的空间模式呈现出在 C3 和 Cz 电极 (CSP1) 以及 C4 电极 (CSP3) 感觉运动皮层区域较强的激活，而 F3 电极左侧额叶区域 (CSP2) 则可能反映了发音过程中语音想象激活的 Broca 区。相反，控制条件的空间模式基本呈现的是由视觉刺激诱发的枕叶和顶叶皮层区域的活动^[32,33]。

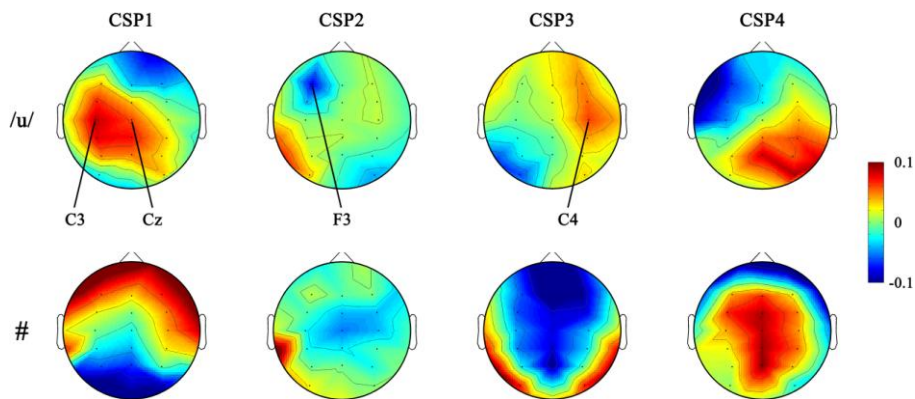


图 5. 音位发音想象和控制条件的 CSP 空域分析结果

3. 4 配对分类结果

九类刺激任务随机选取训练和测试数据集 20 次进行两两配对分类 (共 36 对组合) 的平均分类正确率和标准差如图 6(a)所示。图 6(b)则展示了基于表 1 中发音控制二值描述计算得出的 Jaccard 距离对各刺激材料进行层次聚类的结果。此外，我们还对 36 种组合的分类正确率和各刺激材料间的 Jaccard 距离作了相关性分析，结果显示二者具有很强的相关性($r = 0.58$, $P < 0.001$)。刺激材料间距离越近则表明发音控制相似度越高，进而导致更低的脑电发音想

象分类正确率。音位想象任务和控制条件的组合分类正确率最高，其中最好的分类效果达到了 83% (/u/ vs. #)，这与我们的预期相一致，因为音位发音想象和无想象任务相比，感觉运动皮层区域的激活程度明显不同。而发音控制较相近的音位组合如/u/ vs. /y/的分类正确率则接近基线水平，仅为 51%。

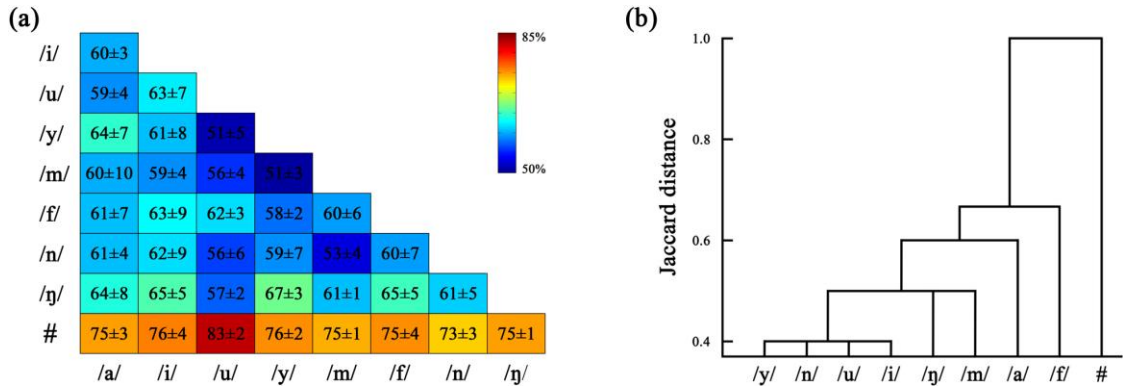


图 6. (a) 配对分类的平均分类正确率和标准差；(b) 刺激材料层次聚类结果

4 总结与讨论

本文提出了一个基于汉语音位想象的脑机接口系统框架，使得受试者使用脑机接口系统时更加自然流畅。三位受试者参与了本研究，实验过程中受试者被要求想象四个汉语元音和四个辅音音位的发音部位及语音发音，同时记录其脑电数据。在数据处理阶段，为了全面深入地分析音位想象脑电数据特征，本文对采集到的脑电信号作了频域分析（ERSP 分析），时域分析（ERP 分析），以及空域分析（CSP 分析），并且提取出音位发音想象效应在频域、时域、空域中最优化的特征向量用于提高脑电数据的分类效果。实验结果表明，音位发音想象任务和控制条件相比具有较高的分类正确率，最高可达 83%，为基于音位发音想象的汉语脑机接口系统研究提供了理论基础。同时，刺激材料间的 Jaccard 距离和分类正确率的高度相关性表明，音位发音想象与运动想象相一致，均可由人脑感觉运动皮层区域的头皮脑电信号来解码预测。

根据前人研究，本研究推测音位发音想象效应的频域特征与运动想象相似，其本质是复杂的发音器官运动想象，故可能会反映 8~12 Hz 的运动控制 μ 节律脑部活动，该频段经常被基于运动想象的脑机接口研究所采用^[34]。另外，由于音位想象任务还包括各音位的发声想象，故其频域特征可能还会反映人脑处理音位发音的频段 2~9 Hz^[35]。脑电实验结果验证了本研究的猜想，ERSP 分析结果表明，音位发音想象任务和控制条件相比在双侧感觉运动皮层区域诱发了 2~10 Hz 频段的频谱能量减弱，这一结果也验证了 Wang 的发现^[35]，差别仅在于他们的研究是针对英语音位的听觉感知，而本实验是关于汉语音位的发音想象。由此也能看出，语音的产生和理解在电生理学层面具有一定的相似性，反映出人脑对语音的输入和产出这两个逆过程可能存在某一部分共享的神经通路。

脑电信号时域分析结果表明，音位发音想象任务和控制条件相比其感觉运动皮层电极信号在时间进程上主要有三个不同阶段，其中最早的差异时段（98~142 ms）被认为是由于视觉刺激物理属性不同而引起的视觉诱发电位（Visual Evoked Potential, VEP），其峰值潜伏期通常在刺激呈现后 90~150 ms，并非由发音想象任务引起，故不属于音位发音想象效应^[32,33]。除去最早的视觉诱发电位差异，本文认为音位发音想象效应主要有两个时间阶段（308~396 ms, 446~530 ms），这一结论与真实发音器官运动引起的颅内及头皮电位时间进程相似^[36,37]，而前人研究亦表明，运动想象相关电位和与其对应的真实运动引发的电生理学信号在形态上非常相似^[38]。需要注意的是，在前人研究中 Dassalla 报告了一个所谓的发音想象语言相关 ERP 成分^[21]，但正如在第 2.2 节中所提到的，该实验设计仅使用了白屏作为控制条件，而没

有选用与发音任务相当的符号作为视觉刺激,因而其所谓的发音想象语言相关电位并没有反映与语言有关的具体脑电特征,其波形差异更多地是由不同实验任务中是否有视觉符号刺激呈现所引起的。

共同空间模式分析结果表明,音位发音想象效应的头皮空间模式分布特征在感觉运动皮层区域显示出较强的脑部活动,尤其是电极 C3 及 C4 标记的双腹侧感觉运动皮层区域(lateral ventral Sensory-Motor Cortex, vSMC), 该区域覆盖了头部发音器官及声带的皮质延髓映射和传入神经支配。总体来看,音位发音想象效应的空间特征是一个双侧化现象,在非优势的大脑右半球依然呈现出较强的脑部活动(如图 5 中音位/u/的 CSP3)。当然,由于处理语言的脑区大多分布在大脑左半球,左侧相比于右侧呈现出更强的脑部活动(如音位/u/的 CSP1)。此外,本研究还发现左前额脑区也呈现出较强的脑部活动,该区域的响应可能是由于左前额负责语音生成的 Broca 区在发声想象任务中被激活而引起的,也可能同语言产生的运动特征相关^[39]。脑电信号的空间模式特征证实,基于语音发音想象的脑机接口系统可被视为语言特定的运动想象脑机接口系统。

基于音位刺激材料和控制条件间 Jaccard 距离的层次聚类分析将 9 类实验刺激材料聚成音位想象任务和无想象任务两个大类,这与两两配对数据分类的结果相一致。音位发音想象任务和控制条件的类别间距离较大,因而在各组配对分类中正确率最高,其中分类效果最好的一组达到了 83% (/u/ vs. #); 而不同音位发音控制的类别内距离相对较小,因而其分类正确率较低,其中发音最相近的一组分类正确率最低,仅为 51% (/u/ vs. /y/) 接近机会水平。配对分类正确率与刺激材料间 Jaccard 距离的高度相关性证实了本文在表 1 中对于各音位发音控制的二值描述是可靠的,该描述不仅反映了音位发音过程的物理距离,也能较好地反映不同音位发音想象脑电信号间的距离,即距离越近则发音控制相似度越高,导致分类正确率越低。同前人实验结果相比,尽管本文选用了—个视觉上与发音想象任务更相近的符号刺激作为控制条件而不是简单的白屏,增加了分类难度,但本实验结果同前人报告的/a/ vs. #及/u/ vs. #组别相比依然提高了受试间平均分类正确率。由此可见,对音位发音想象脑电数据进行频域和时域优化,提取特征时增加最优的频域和时域特征能够显著提高音位发音想象脑电数据分类正确率。此项研究验证了本文的假设,即音位发音想象任务可被视为复杂的发音器官运动想象任务,并且发音运动想象头皮脑电信号可作为系统开发应用研究的基础,控制无创脑机接口系统作为语言神经假体来实现人脑同外部机器设备之间的直接沟通交流。

本项研究对于开发基于语音想象的脑机接口系统及临床康复有重要指导意义。BCI 技术和基于 BCI 技术的临床康复研究都处于初级阶段,难免问题重重,但是随着研究的深入,一个个问题终将被解决。其发展前景是模块化和集成化,模块化是集成化的基础,集成化是推广使用的前提。识别率高、信号处理速度快、采集功能通用化、算法处理功能系统化、适应性强、合理引入反馈、评判标准化的 BCI 技术有待进一步研究。处理多种或者大部分信号,机构设计绿色且实用、响应迅速、控制精准、运动模式多样、评价系统科学的临床康复技术也有待进一步研究。如果模块化的脑机接口系统和临床康复技术之间能够合理地柔性组合,那么 BCI 研究将迈入理论与临床应用相结合的崭新阶段。基于语音想象的脑机接口系统研究将使得 BCI 系统的控制操作更人性化、更接近人类言语交流的自然状态,同时也将对语言加工认知神经机制的研究提供新的研究视角和思路。

参考文献

- [1] Wolpaw, J. R., N. Birbaumer, D. J. McFarland, et al. Brain-Computer Interfaces for Communication and Control [J]. *Clinical Neurophysiology*, 2002, 113(6): 767-791.
- [2] Wolpaw, J. R., N. Birbaumer, W. J. Heetderks, et al. Brain-Computer Interface Technology: A Review of the First International Meeting [J]. *IEEE Transactions on Rehabilitation Engineering*, 2000, 8(2): 164-173.
- [3] He, B., S. Gao, H. Yuan, et al. Brain Computer Interface [C]//Bin He. *Neural Engineering, 2nd ed.* US: Springer, 2013: 87-151.
- [4] Hochberg, L. R., M. D. Serruya, G. M. Friehs, et al. Neuronal Ensemble Control of Prosthetic Devices by a Human with Tetraplegia [J]. *Nature*, 2006, 442(7099): 164-171.
- [5] Kennedy, P. R., M. T. Kirby, M. M. Moore, et al. Computer Control Using Human Intracortical Local Field Potentials [J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2004, 12(3): 339-344.
- [6] Leuthardt, E. C., G. Schalk, J. R. Wolpaw, et al. A Brain-Computer Interface Using Electrocorticographic Signals in Humans [J]. *Journal of Neural Engineering*, 2004, 1(2): 63-71.
- [7] Birbaumer, N., N. Ghanayim, T. Hinterberger, et al. A Spelling Device for the Paralyzed [J]. *Nature*, 1999, 398(6725): 297-298.
- [8] Cheng, M., X. Gao, S. Gao, et al. Design and Implementation of a Brain-Computer Interface with High Transfer Rates [J]. *IEEE Transactions on Biomedical Engineering*, 2002, 49(10): 1181-1186.
- [9] Donchin, E., K. M. Spencer and R. Wijesinghe. The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface [J]. *IEEE Transactions on Rehabilitation Engineering*, 2000, 8(2): 174-179.
- [10] Wolpaw, J. R., D. J. McFarland and T. M. Vaughan. Brain-Computer Interface Research at the Wadsworth Center [J]. *IEEE Transactions on Rehabilitation Engineering*, 2000, 8(2): 222-226.
- [11] Comerchero, M. D. and J. Polich. P3a and P3b From Typical Auditory and Visual Stimuli [J]. *Clinical Neurophysiology*, 1999, 110(1): 24-30.
- [12] Jansen, B. H., A. Allam, P. Kota, et al. An Exploratory Study of Factors Affecting Single Trial P300 Detection [J]. *IEEE Transactions on Biomedical Engineering*, 2004, 51(6): 975-978.
- [13] Croft, R. J., C. J. Gonsalvez, C. Gabriel, et al. Target-to-Target Interval Versus Probability Effects on P300 in One and Two-Tone Tasks [J]. *Psychophysiology*, 2003, 40(3): 322-328.
- [14] Polich, J., and M. R. D. Heine. P300 Topography and Modality Effects from a Single-Stimulus Paradigm [J]. *Psychophysiology*, 1996, 33(6): 747-752.
- [15] Galan, F., M. Nuttin, E. Lew, et al. A Brain-Actuated Wheelchair: Asynchronous and Non-invasive Brain-Computer Interfaces for Continuous Control of Robots [J]. *Clinical Neurophysiology*, 2008, 119(9): 2159-2169.
- [16] McFarland, D. J., W. A. Sarnacki and J. R. Wolpaw. Electroencephalographic (EEG) Control of Three-dimensional Movement [J]. *Journal of Neural Engineering*, 2010, 7(3): 036007.
- [17] Wolpaw, J. R. and D. J. McFarland. Control of a Two-dimensional Movement Signal by a Noninvasive Brain-Computer Interface in Humans [J]. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 2004, 101(51): 17849-17854.
- [18] Blankertz, B., G. Dornhege, C. Schafer. Boosting Bit Rates and Error Detection for the Classification of Fast-paced Motor Commands Based on Single Trial EEG Analysis [J]. *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, 2003, 11(2): 127-130.
- [19] Oldfield, R. C. The Assessment and Analysis of Handedness: the Edinburgh Inventory [J]. *Neuropsychologia*, 1971, 9(1): 97-113.
- [20] Crystal, D. Linguistics [M]. Baltimore: Penguin Books, 1971.
- [21] DaSalla, C. S., H. Kambara, M. Sato, et al. Single-trial Classification of Vowel Speech Imagery Using Common Spatial Patterns [J]. *Neural Networks*, 2009, 22(9): 1334-1339.
- [22] Delorme, A. and S. Makeig. EEGLAB: An Open Source Toolbox for Analysis of Single-trial EEG Dynamics Including Independent Component Analysis. [J]. *Journal of Neuroscience Methods*, 2004, 134(1): 9-21.
- [23] Makeig, S., S. Debener, J. Onton, et al. Mining Event-related Brain Dynamics [J]. *Trends in Cognitive Sciences*, 2004, 8(5): 204-210.
- [24] Ramoser, H., J. Muller-Gerking and G. Pfurtscheller. Optimal Spatial Filtering of Single Trial EEG during Imagined Hand Movement [J]. *IEEE Transactions on Rehabilitation Engineering*, 2000, 8(4): 441-446.
- [25] Dornhege, G., B. Blankertz, G. Curio, et al. Boosting Bit Rates in Noninvasive EEG Single-Trial Classifications by Feature Combination and Multiclass Paradigms [J]. *IEEE Transactions on Biomedical Engineering*, 2004, 51(6): 993-1002.
- [26] Grosse-Wentrup, M. and M. Buss. Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction [J]. *IEEE Transactions on Biomedical Engineering*, 2008, 55(8): 1991-2000.
- [27] Lotte, F., M. Congedo, A. Lecuyer, et al. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces [J]. *Journal of Neural Engineering*, 2007, 4.
- [28] Burges, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition [J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121-167.
- [29] Weston, J. and C. Watkins. Multi-class Support Vector Machines. Technical Report CSD-TR-98-04 [R]. Royal Holloway University of London: Department of Computer Science, 1998.
- [30] Chang, C. C. and C. J. Lin. LIBSVM: A Library for Support Vector Machines [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, 2(3): 27.
- [31] Hsu, C. W., C. C. Chang and C. J. Lin. A Practical Guide to Support Vector Classification. 2003.
- [32] DiRusso, F., A. Martinez, M. I. Sereno, et al. Cortical Sources of the Early Components of the Visual Evoked

Potential [J]. *Human Brain Mapping*, 2002, 15(2): 95-111.

[33] Hillyard, S. A. and L. Anllo-Vento. Event-related Brain Potentials in the Study of Visual Selective Attention [J]. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 1998, 95(3): 781-787.

[34] McFarland, D. J., L. A. Miner, T. M. Vaughan, et al. Mu and Beta Rhythm Topographies during Motor Imagery and Actual Movements [J]. *Brain Topography*, 2000, 12(3): 177-186.

[35] Wang, R., M. Perreau-Guimaraes, C. Carvalhaes, et al. Using Phase to Recognize English Phonemes and Their Distinctive Features in the Brain [J]. *Proceedings of the National Academy of Sciences (PNAS)*, 2012, 109(50): 20685-20690.

[36] Deecke, L., M. Engel, W. Lang, et al. Bereitschaftspotential Preceding Speech after Holding Breath [J]. *Experimental Brain Research*, 1986, 65(1): 219-223.

[37] Wohlert, A. Event-Related Brain Potentials Preceding Speech and Nonspeech Oral Movements of Varying Complexity [J]. *Journal of Speech and Hearing Research*, 1993, 36(5): 897-905.

[38] do Nascimento, O. F., K. D. Nielsen and M. Voigt. Movement-Related Parameters Modulate Cortical Activity during Imaginary Isometric Plantar-Flexions [J]. *Experimental Brain Research*, 2006, 171(1): 78-90.

[39] Lazar, R. M., R. S. Marshall, J. Pile-Spellman, et al. Interhemispheric Transfer of Language in Patients with Left Frontal Cerebral Arteriovenous Malformation [J]. *Neuropsychologia*, 2000, 38(10): 1325-1332.

作者简介:

杨晓芳 (出生于 1989 年), 女, 硕士, 主要研究领域为认知神经科学、生物医学工程。

本文通讯作者, Email: xfyang.thu@gmail.com



江铭虎 (出生于 1962 年), 男, 教授, 主要研究领域为自然语言处理、认知科学。

Email: jiang.mh@tsinghua.edu.cn

