



基于耳周肌电信号的默念口令识别方法

魏柏淳, 姜峰, 张松涛, 张琦, 段锦楠, 王修来

引用本文:

魏柏淳, 姜峰, 张松涛, 等. 基于耳周肌电信号的默念口令识别方法[J]. *智能系统学报*, 2025, 20(4): 894-904.
WEI Baichun, JIANG Feng, ZHANG Songtao, et al. Method for silent command recognition based on periauricular EMG signals[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(4): 894-904.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202406017>

您可能感兴趣的其他文章

一致性协议匹配的跨模态图像文本检索方法

Matching with agreement for cross-modal image-text retrieval

智能系统学报. 2021, 16(6): 1143-1150 <https://dx.doi.org/10.11992/tis.202108013>

改进Faster R-CNN的汽车仪表指针实时检测

Improved Faster R-CNN vehicle instrument pointer real-time detection algorithm

智能系统学报. 2021, 16(6): 1056-1063 <https://dx.doi.org/10.11992/tis.202011003>

面向车规级芯片的对象检测模型优化方法

Object detection model optimization method for car-level chips

智能系统学报. 2021, 16(5): 900-907 <https://dx.doi.org/10.11992/tis.202107057>

基于卷积神经网络的盲文音乐识别研究

Research on braille music recognition based on convolutional neural networks

智能系统学报. 2019, 14(1): 186-193 <https://dx.doi.org/10.11992/tis.201805002>

基于改进型BP神经网络的手部动作识别

Hand-motion recognition based on improved BP neural network

智能系统学报. 2018, 13(5): 848-854 <https://dx.doi.org/10.11992/tis.201703018>

基于卷积特征和贝叶斯分类器的人脸识别

Face recognition based on convolution feature and Bayes classifier

智能系统学报. 2018, 13(5): 769-775 <https://dx.doi.org/10.11992/tis.201706052>

DOI: 10.11992/tis.202406017

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20250408.0943.002>

基于耳周肌电信号的默念口令识别方法

魏柏淳¹, 姜峰², 张松涛¹, 张琦¹, 段锦楠¹, 王修来³

(1. 哈尔滨工业大学生命科学与医学学部, 黑龙江哈尔滨 150000; 2. 南京信息工程大学未来技术学院, 江苏南京 211800; 3. 东部战区总医院, 江苏南京 210018)

摘要: 智能设备的普及促使可穿戴人机交互技术需求日益增加。为提高用户接受度, 人机交互技术对交互易用性与隐蔽性要求较高。本文提出基于耳周肌电信号的默念口令识别方法。该方法易于与集成生理电采集的耳机设备结合, 实现无声操控智能设备, 减少社交尴尬。具体地, 本文首先确定并构建口令经验原则, 筛选最优口令集。其次, 根据单通道信噪比和分类准确率选择最优耳周传感器位置。再次, 提出基于 CNN-Transformer 结构的识别模型构建耳周肌电信号与默念口令的时空映射。最后, 大量实验评估方法可行性和稳定性。结果表明, 本文方法平均准确率 91.18%, 优于相关任务的先进模型, 且在命令变形和头部运动下表现稳定。本文方法奠定了默念口令识别商业产品的技术基础。

关键词: 人工智能; 模式识别; 人机交互; 神经人接口; 人体意图解码; 默念口令识别; 肌电信号处理; 神经网络
中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2025)04-0894-11

中文引用格式: 魏柏淳, 姜峰, 张松涛, 等. 基于耳周肌电信号的默念口令识别方法 [J]. 智能系统学报, 2025, 20(4): 894-904.

英文引用格式: WEI Baichun, JIANG Feng, ZHANG Songtao, et al. Method for silent command recognition based on periauricular EMG signals[J]. CAAI transactions on intelligent systems, 2025, 20(4): 894-904.

Method for silent command recognition based on periauricular EMG signals

WEI Baichun¹, JIANG Feng², ZHANG Songtao¹, ZHANG Qi¹, DUAN Jinnan¹, WANG Xiulai³

(1. Department of Life Science and Medicine, Harbin Institute of Technology, Harbin 150000, China; 2. Nanjing University of Information Science and Technology, School of Future Technology, Nanjing 211800, China; 3. General Hospital of Eastern Theater, Nanjing 210018, China)

Abstract: The widespread use of smart devices has led to an increasing demand for wearable human-computer interaction technologies. To improve user acceptance, human-computer interaction technologies require high levels of interaction usability and concealment. This paper proposes a method for silent command recognition based on periauricular EMG signals. This method is easy to integrate with headphones equipped with integrated physiological signal acquisition, enables silent control of smart devices, and reduces social awkwardness. First, the command empirical principles are determined and constructed, and then the optimal command set is selected through screening. Second, the optimal periauricular sensor positions are chosen based on single-channel signal-to-noise ratio and classification accuracy. Third, a recognition model based on the CNN-Transformer structure is proposed to learn the spatiotemporal mapping between periauricular EMG signals and silent commands. Finally, extensive experiments evaluate the feasibility and stability of this method. Results demonstrate that the average accuracy of this method is 91.18%. The proposed method is superior to advanced models in similar tasks and is stable under command deformation and head motion. This method lays the technical foundation for commercial products of silent command recognition.

Keywords: artificial intelligence; pattern recognition; human-computer interaction; neural human-machine interface; human intent decoding; silent command recognition; EMG processing; neural networks

可穿戴人机交互技术因智能手表、耳机或眼镜等智能设备的普及和易于操控的特点而日益受用户欢迎^[1-4]。这些交互的日常应用, 比如开车时接听电话或运动时切换歌曲, 需要一种“微交互”的方式, 即吸引用户较少注意力且不会干扰用户主要任务的交互方式。基于默念口令的人机交互

大大满足了这一要求, 具有无需使用手势与语音交互、极大降低公共场所社交尴尬等优势。

面部动作识别是基于默念口令的人机交互领域主要技术之一。传统基于视觉的方法可以识别并跟踪面部动作^[5-6], 甚至可以识别用户情绪^[7-8]。尽管行之有效, 但这类方法需要摄像头持续处于用户面部前方, 可能会受到遮挡或光照条件的影响, 且存在社交尴尬等情况。以上缺陷限制了此类方法在室外和光照受限环境以及非固定条件下

收稿日期: 2024-06-11. 网络出版日期: 2025-04-08.

基金项目: 江苏省科技计划项目 (BE2021086); 中央引导地方科技发展专项项目 (2024ZYD0266).

通信作者: 王修来. E-mail: wangxiulai@126.com.

的日常使用。头戴式摄像头^[9]可能会缓解这些问题,但过高的成本与社交尴尬仍然限制了该类技术的实际应用。基于音频的方法可以通过可穿戴麦克风识别语音命令^[10-11],这可能会解决遮挡和设备成本问题,然而,这类技术较难应用于嘈杂的公共环境或健身房与游泳池等交互场景。

现有面向面部动作和默念语音识别的可穿戴设备提供了替代方案,与默念口令识别任务不同,现有主流技术使用磁性^[12]、电容^[13]、眼电图^[14]或表面肌电信号(surface electromyography, EMG)^[15-17]来识别离散的面部动作。然而,在微交互的背景下,大多数面部动作往往都较为夸张,且缺乏面部动作与相应操作之间的语义联系(比如默念“停止”或“切换歌曲”口令来操控耳机),因此交互过程较不自然。此外,用于持续跟踪面部动作的方法无法区分较为相似的面部动作,因此不适用于默念口令识别任务。默念语音识别(silent speech recognition, SSR)旨在利用表面EMG^[18-21]、脑电图^[20, 22]或超声波传感器^[23]转换和回归默念沟通,以应用于患有声音或言语障碍的患者。该任务同样与默念口令识别研究差异较大。此外,SSR传感器通常覆盖颊部、下颚甚至舌头^[24]。尽管这些传感器位置可以跟踪控制言语表达的肌肉运动,但它限制了健康消费者的日常使用和接受度。据目前所知,尚无相关方法能够实现可穿戴、隐蔽和微交互需求。

本文通过利用耳周EMG信号识别默念语义命令来实现这一目标,该方法具有不引人注目、社交尴尬最小化和可靠性高等特点。尽管当前的研究已经证明了使用面部EMG传感器来检测和跟踪面部动作和默念语音的可行性,但仍然存在一些关键挑战需要解决。

1) 耳周传感。为了提高隐蔽性和用户接受度,生理信号应尽量远离面部,以便吸引更多广泛的消费群体。因此,传感器安放在耳周附近是一个较理想的选择。然而,由于用于控制嘴部的肌肉很少位于耳周,耳周传感通常会从面颊、下颌或舌头的肌肉中收集到串扰信号,因此不可避免地会降低信号的质量和信噪。

2) 默念口令识别。自然口令会导致相对较小且相似的口型、面部动作和肌肉运动,这可能会降低口令的可分性。因此,口令的筛选和识别模型的设计对于建立耳周EMG信号和默念口令之间的时空映射至关重要。

3) 真实环境适配。本文从以下方面提高和评估默念口令识别方法的实用性:①命令变形,用

户默念口令时添加拟声词,例如将口令“停止”执行为“停止吧”或“停止啊”,模型仍应准确识别口令。此外,模型应适应不同用户执行口令的节奏与幅度。②非静止条件,如果用户在行走或移动头部时,口令识别模型性能不应显著下降。③不同的传感协议,系统应适应不同的传感采集设置,并提高算法对信号采集系统的兼容性。

为了克服以上挑战,本研究探索了通过耳周传感器识别静默口令的可行性。具体地,首先通过设置耳周传感通道来限制传感器布置空间,以便与耳机集成并提高隐蔽性。其次,根据信号分析和口令语义考虑选择最优传感器布置位置和口集。再次,通过采集的EMG信号训练模型,并将其与其他用于相关任务的先进模型进行比较。最后,在命令变形、非静止条件以及不同空间和时间分辨率信号等条件下评估模型的准确性和可靠性。本文的关键贡献总结如下:

1) 首次提出基于耳周EMG信号的默念口令识别方法,进而提升基于默念口令人机交互的隐蔽性与易用性;

2) 基于实际应用、信号分析与实验结果综合考虑,确定默念口令识别框架下最优传感器布置方案与默念口令集;

3) 提出基于CNN-Transformer结构的模型以建立EMG信号与默念口令之间的映射,并将其与用于相关任务的先进神经网络模型比较;

4) 数据采集过程招募45名受试者执行10种默念口令,并针对真实场景进行广泛的实验以评估本方法的兼容性与鲁棒性。

1 相关工作

1.1 面向面部EMG解码的可穿戴设备设计

本研究从交互隐蔽性和信号质量的角度考虑可穿戴传感器配置,即如何更好地定位唇部肌肉并最大限度地减少社交尴尬。基于面部肌电信号解码的可穿戴传感设备通常基于发音器官的解剖结构,选择将EMG传感器放置在面颊和下颌处。相关研究中的可穿戴设备设计如图1所示,Deng等^[25]和Wand等^[26]分别放置8通道与5通道EMG传感器于面颊和下颌的相关发声肌肉上,并实现了约10%的单词识别错误率。尽管此类传感设备实现了较高的口令识别性能,但违背了交互的自然性与非侵入性要求,可能降低用户的接受度。Matthies等^[27]探索基于耳内电生理场传感的面部动作识别框架,并针对5种口令达到了约90%的识别准确率。然而如文献^[28]所述,

耳内传感器主要检测 EEG、EOG 和 EMG 混合信号,可能会降低所收集信号的信噪比。Wu 等^[29]探索了面部与耳部 EMG 组合式电极布置方案,并开发了用于面部动作跟踪的单耳侧肌电传感系统。不同于已有研究,本文面向人机交互自然性与隐私性需求,评估并选择耳周 EMG 传感通道配置,进而构建耳周肌电信号与语义式默念口令之间的映射,进而支撑耳机集成式人机交互技术开发。



图 1 相关研究中的可穿戴设备设计

Fig. 1 Design of wearable devices in related research

1.2 基于 EMG 信号的默念语音识别

学习耳周 EMG 信号和默念口令之间时空映射的识别模型在本文整体技术框架中起关键作用,并可借鉴类似任务的模型结构。SSR 模型通过分类音素和词语来连续跟踪受试者意图表达的语句。Meltzner 等^[25]提出多阶段框架,并使用隐马尔可夫模型和上下文关联的三音素分类模型。Kim 等^[30]提出通过长短期记忆网络 (long short-term memory, LSTM) 建模 EMG 信号相邻时间窗之间的时间连接。以上方法主要关注人工 EMG 特征提取并建模特征之间的时间关系,以回归面

部肌肉运动并将其转换为单词和句子。不同于以上面向语句预测的回归方法,本方法主要关注离散的默念口令识别,并研究 EMG 通道之间的空间依赖性与端到端神经网络模式识别模型。

1.3 基于 EMG 信号的面部动作识别

基于 EMG 信号的面部动作识别 (facial gesture recognition, FGR) 关注于跟踪面部活动,其通常涉及不同的面部肌肉激活模式。Wu 等^[29]采集 6 通道面部与耳周 EMG 信号,并利用卷积神经网络回归通过视频信号得到的二维面部坐标点,达到了约 1.85 mm 的回归精度误差。不同于 FGR,在默念口令识别任务中许多口令的肌肉激活模式相似,因此不同口令下的肌电信号分布具有一定的重合性。此外,本研究的耳周传感配置进一步增加了默念口令识别的技术难度。在口令执行时,耳周的肌肉很少被激活,因此主要依赖于发音肌肉的串扰信号。为了克服以上挑战,本文研究端到端建模不同通道与时间窗之间的时空关系。

2 耳周传感默念口令识别框架

本文技术框架如图 2 所示,首先受试者根据交互界面引导执行默念口令,在此过程中收集受试者 EMG 信号与对应标签。随后,通过收集得到的数据集训练口令识别模型,最后在测试阶段验证该模型。具体地,首先通过实验探索筛选最优口令集。随后构建训练集,并训练基于 CNN-Transformer 的默念口令识别模型。最后,基于测试集验证模型在理想与非理想场景下的识别性能。

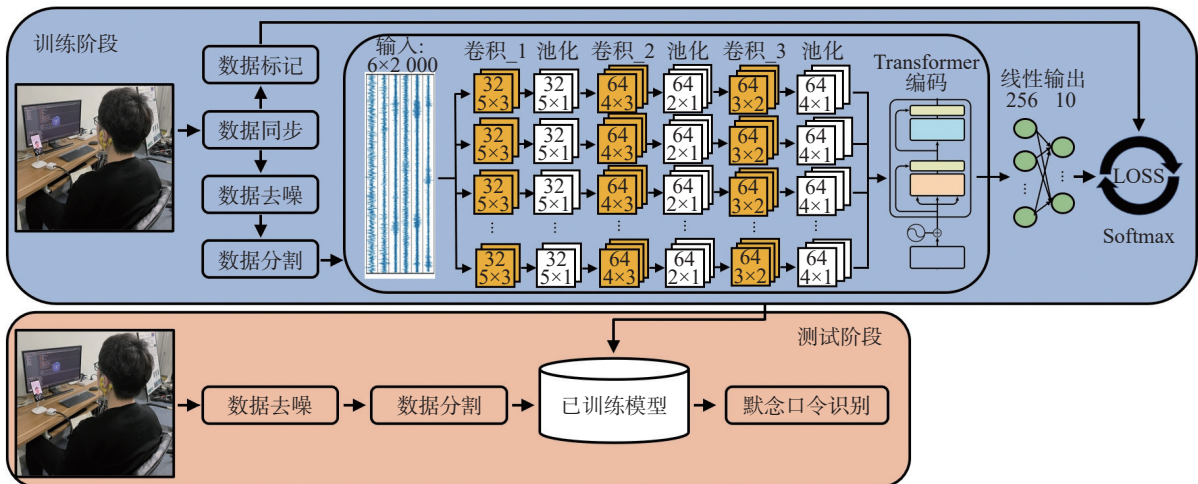


图 2 默念口令识别技术框架

Fig. 2 Framework of silent command recognition

2.1 默念口令筛选

耳周传感器配置与任务需求会显著影响口令的可分离性。此外, 其他因素例如不同口令间口型和肌肉激活的潜在差异同样影响口令的可分性。因此, 本节重点研究默念口令集的构建。为此, 首先通过将常见音素分成呈现相似模式和分布的子组并分析其 EMG 信号。随后, 根据分析准则确定最优口令。尽管最优口令选择和传感器配置之间可能存在交互作用, 但本文认为口令选择是影响最终性能的更为重要因素, 因此在此步骤中使用完备传感器集以提供尽可能多的信息。这里, 使用如图 3 所示的初始传感器配置。

实验方案 本实验招募 10 名受试者。每个受试者坐在屏幕前, 并根据屏幕上提示默念 19 个音素(数据集构建细节在第 2.3 节中描述)。聚类结果如图 4 所示, 本节根据口型类别选择代表性元音和辅音音素, 包括大口型 (a, o)、半开口型 (u,

e, ü, i)、爆破音 (b, d, g)、卷舌音 (zh, er) 和非卷舌音 (z, h, l, f, n, m, ng, j)。音素以随机顺序提示, 每个音素执行 20 次, 不用音素之间有 2~5 s 休息时间。为了获取标签, 每个受试者在默读音素时按下标记按钮。由于数据标记设备共有 5 个按钮, 因此每个受试者的实验分为 4 个阶段以覆盖所有 19 个音素。各阶段之间有 1~3 min 的休息以避免疲劳。

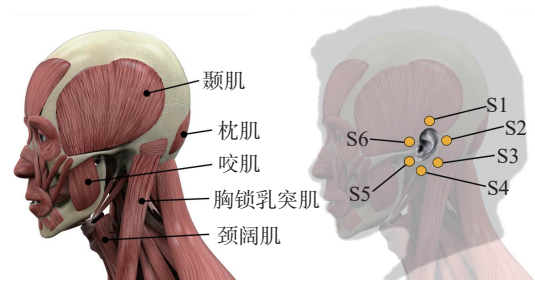


图 3 潜在相关肌肉和传感器配置
Fig. 3 Potential related muscles and sensor configurations

	超过 80% 样本位于该类别			60%~80% 样本位于该类别			40%~60% 样本位于该类别												
音素	b	d	g	z	h	l	f	n	m	ng	j	zh	er	a	o	u	e	ü	i
类别 1	0	0	21	20	7	32	0	0	0	10	26	44	1	40	43	6	47	1	26
类别 2	100	100	0	0	0	0	100	30	100	0	0	0	1	0	0	0	0	0	0
类别 3	0	0	42	16	68	1	0	65	0	0	18	4	96	0	15	5	5	65	18
类别 4	0	0	0	62	0	0	0	0	0	0	51	0	0	41	0	0	0	0	51
类别 5	0	0	37	2	25	67	0	5	0	90	5	52	2	19	42	89	48	34	5

图 4 音素聚类结果
Fig. 4 Clustering results of the phonemes

分析方法 结果分析阶段使用经典 EMG 特征集, 即 Hudgin’s 特征集^[31] 和 Du’s 特征集^[32] 对时间窗数据进行表达。在比较了不同时间窗长度 (100~2000 ms) 下的音素分类准确性后, 最终确定时间窗长度为 2000 ms, 步进为 150 ms。

首先, 本节基于 K-means 聚类分析音素间可分性, 并使用每个音素类别在一个簇中样本所占其总样本数量的比例来评估聚类结果。簇的数量范围从 3 到 10, 并根据音素最大平均比例确定 5 为最优簇数量。如图 4 所示, 百分比表示每个音素类别在簇中的样本比例, 并用颜色强调比例超过 40% 的音素。其中, 同一簇中的音素表示它们具有相似的分布, 也意味着可能难以通过耳周 EMG 信号区分同一簇中的不同音素。

为了进一步证明, 本节针对每个簇中的百分比数值前 4 位的音素类别进行 4 分类任务, 并应用随机森林作为分类器。表 1 给出了各簇内前 4 百分比音素类别的分类准确率。从表格中可以看出, 同一簇中不同音素之间的可分性较差。

表 1 同一簇内不同音素间分类准确率
Table 1 Classification accuracy between different phonemes within the same cluster %

簇类别	音素	分类准确率
类别 1	e, zh, a, o	66.80
类别 2	b, m, f, d	47.17
类别 3	er, h, n, ü	81.21
类别 4	z, i, j, a	72.00
类别 5	ng, u, l, zh	53.71

随后, 本节分析音素之间的整合方式。在这里, 随机选择 5 个因素构建 5 分类任务, 并重复 2000 次。根据分类准确率排序, 图 5 给出了最高可分性的 10 个音素组合与最低可分性的 10 个音素组合, 该结果进一步证明了聚类分析的结果, 即同一簇内的音素可分性远低于不同簇间音素的可分性。此外, 基于分类结果可得到以下规律:

1) 音素“b(p), m, f, d(t)”之间很难区分, 但如果其中一个音素与其他音素构建组合, 则可分性较

高。这意味着该簇的分布具有较小的类内距离，而与其他之间的类间距离则相对较大；

2) 与其他音素相比，音素“ü”和“er”在前 5 和前 10 的音素组合中出现的次数最多；

3) 音素“z(c, s), i, j(q, x)”较难区分。因此，组合内应最多存在这几种音素中的一种；

3) 音素“ng, u, l, zh, e, o”较难区分。因此，组合内应最多存在这几种音素中的一种。

音素					准确率
g	er	ü	f	i	95.54%
b	ng	ü	er	z	95.49%
ü	er	z	d	a	95.36%
ng	er	m	d	a	95.19%
er	b	a	z	ü	95.16%
er	ü	l	i	b	95.00%
ü	a	d	er	j	94.99%
f	er	l	ü	z	94.87%
er	ü	a	g	f	94.81%
zh	z	ü	er	b	94.61%

(a) 最高准确率 10 种音素组合

音素					准确率
m	n	b	zh	l	60.39%
o	z	l	zh	e	60.36%
n	l	o	g	h	60.29%
n	f	zh	l	d	59.76%
h	e	l	g	ng	59.49%
d	f	b	m	ü	58.77%
m	d	z	b	f	57.44%
m	zh	b	d	f	56.98%
d	o	m	f	b	56.31%
h	n	f	d	g	56.11%

(b) 最低准确率 10 种音素组合

图 5 前 10 位和后 10 位音素组合准确率

Fig. 5 Accuracy of the top ten and bottom ten phoneme combinations

口令选择与口令集构建 根据音素分析结果作为先验，并结合口令语义构建如表 2 所示的口令集合。具体地，通过将可区分的音素包含在第一个汉字中以构建口令，随后通过考虑语义上下文扩展音素。此实验招募 5 名受试者，并使用初始传感器配置和与前一部分相同的实验范式收集数据。在分析阶段，同样进行聚类、簇内分类和随机类别组合分类任务，以评估口令的可分离性并筛选出可分性最高的前 10 个口令。具体地，将口令样本聚类为 10 个簇，分类同一簇内的前 4 口令类别，最后从整个口令集中随机选择并分类 10 个命令。分析后，确定了以下 10 个命令：快进、开始、上一首、中断、调小、播放、耳机、巨大、允许、履行。

表 2 完备默念口令集

Table 2 Complete silent command set

语义上下文	对应口令类别
开始	开始、播放、打开、继续、执行、履行、允许
结束	停止、关闭、退出、终止、删除、中断、暂停、结束、静音
前进	下一首、快进、调大、音量加、循环、巨大
后退	上一首、退回、调小、音量减、返回、耳机

2.2 传感器布置方案

如图 3 左侧给出的可能被 EMG 传感器检测到的耳周肌肉。进而确定了初始的传感器配置，以便尽可能覆盖或靠近耳周肌肉并避免相邻传感器之间的信号串扰。如图 3 右侧所示，初始传感器配置覆盖了颞肌(S1)、乳突骨(S2)、胸锁乳突肌近邻的耳下(S3)、耳朵底部与下颌骨边缘(S4)、下颌骨和颞骨交界处且与咀嚼肌相近(S5)以及颞肌边缘和颞骨上方(S6)的位置。本研究所使用的 EMG 传感系统包括 3 种类型通道，即测量、参考和接地。由于其附着耳周的骨特性，在左右耳两侧使用 S2 作为参考和地面通道。为了确定最优传感器安放位置，本研究分析了各通道的信噪比(signal-to-noise ratio, SNR)和分类准确率，并同时考虑了电极干扰和佩戴舒适性。

SNR 分析 5 名受试者执行 10 种默念口令，并分析其双侧 12 通道 EMG 信号。通过计算每个口令和每个通道 SNR，并将 SNR 在受试者、口令和左右耳维度上对结果进行平均，得到结果如表 3 所示。其中，S3、S4 与 S5 通道 SNR 最高，而 S1 的 SNR 为负值。由于 S2 作为参考和接地通道，因此不计算其 SNR。本研究 SNR 低于文献 [29] 中所示的结果，这表明基于耳周传感的默念口令识别任务难度更高。此外，没有明显的规律表明哪一侧耳朵具有更好的 SNR。因此，本研究保留两侧耳周 S3、S4 与 S5 通道。

表 3 不同 EMG 通道间 SNR 与准确率对比

Table 3 Comparison of SNR and accuracy between different EMG channels

通道序号	SNR/dB	分类准确率/%
S1	-0.448 6	19.63
S2	—	—
S3	1.794 2	46.53
S4	2.219 9	45.99
S5	1.536 6	53.62
S6	0.443 6	14.67

准确率分析 与默念口令筛选阶段类似，本部分同样使用 Hudgin’s 与 Du’s 特征集以及随机

森林分类器。这里使用每个通道的信号来分别对 10 种口令进行分类。通过使用每个受试者的数据训练和测试分类器, 显示了跨受试者和口令的平均准确率。如表 3 所示, S5 具有最高的平均准确率, 而 S1 和 S6 的准确率相对最低。分类结果与 SNR 结果相符合, 且左右耳未呈现明显的性能差异, 这表明应该保留两侧耳周通道。综合各通道的 SNR 和准确率结果, 决定将 S5、S4、S3 和 S2 作为最终的传感器布置方案。

2.3 数据集构建

本研究招募 45 名受试者(年龄 20~42 岁, 女性 18 名, 男性 27 名)构建耳周传感默念口令数据集。每名受试者被要求坐在屏幕前, 根据屏幕上呈现的提示执行默念命令。每名受试者进行了 4 次试验, 包括 2 次共 10 个无声口令类型的试验(每个口令重复 20 次), 2 min/次静息状态(即保持静默), 以及 2 min/次交谈或朗读试验。通过这种方式, 共获得了 12 种标签类型, 包括 10 个无声口令、静息状态和说话状态。对于 2 次无声口令试验, 口令的顺序随机生成的, 并在命令之间插入 2 s 的间隔。此外, 要求受试者在执行命令时同步按下对应按钮以标记数据。在实验过程中, 不限制口令的执行速度与口型。对于静息状态或说话状态试验, 限制受试者头部或颈部大幅度运动, 但允许小幅度的头部动作或手部动作。

如图 6 所示, EMG 采集系统由 ADS1299 生物放大电路即 OpenBCI、Ag/AgCl 湿电极和蓝牙模块组成。通过这种方式, 各通道收集的信号可放大并由模拟转换为数字信号, 进而滤波并传输到计算机, 最终由算法处理或存储以供使用。



图 6 实验装置示意

Fig. 6 Schematic diagram of experimental setup

本研究将硬件集成到自制 3D 打印盒中, 并将 4 个传感器附在耳周, 对应于 S2、S3、S4 和 S5 位置。EMG 采样频率为 1 000 Hz。为了获得标签, 5 个按钮与 NI myRIO R3670 连接, 频率为 500 Hz。在默念口令数据采集实验中, 受试者按下每个按钮以标记对应口令。按压信号和 EMG

信号通过由 myRIO 发送的高电平作为标记进行同步。所有受试者均为普通话使用者, 轻微口音并不会影响其默念口令的执行。

2.4 肌电数据预处理与默念口令识别模型

数据预处理 在数据同步之后, 应用截止频率为 20~500 Hz 带通滤波器, 以消除 EEG、EoG 及其他噪声成分。此外, 应用频率为 50 Hz 的陷波滤波器以滤除工频噪声干扰。随后, 采用长度 2 000 ms、步进 150 ms 的滑动窗口对数据流进行分段。步进参数设置时考虑了实时处理与数据量要求。标签根据按钮的按压信号生成, 并使用窗口中的主要成分标签作为该窗口的对应标签。在训练阶段, 舍弃命令之间静息状态标签的窗口以降低训练数据噪声。

CNN-Transformer 网络 本文结合 CNN 与 Transformer 结构对 EMG 信号进行时空建模进而提取稀疏表示特征。受文献 [33] 所提出的基于 EMG 的手势识别模型结构启发, 本文采用多流 CNN 网络以从信号中提取多尺度特征, 并使用 2 维卷积核提取局部时间特征和相邻通道的空间依赖性。如图 2 所示, CNN 网络中有 10 个并行处理流, 每个流由 3 个卷积层组成, 每个卷积层后连接 ReLU 激活函数、批归一化层和最大池化层。借鉴于在 EMG 处理领域中 Transformer 网络的成功应用, 在 EMG 数据经过多流 CNN 特征提取后采用 Transformer 编码器对其进行进一步编码以提取全局特征信息 [34]。编码器层由两个子结构组成: 多头自注意力结构与位置前馈结构, 两种结构有序堆叠。最终, 处理后的特征经由两层全连接网络后输出分类结果, 全连接层之间同样由 ReLU 函数相连。训练过程中采用交叉熵函数作为损失函数, 并利用 Adam 优化器更新模型参数, 批数量设置为 128, 循环数为 30 epoch。

3 实验结果与分析

3.1 实验验证与结果

本研究通过筛选语义明确且易于识别的命令, 根据信号质量选择传感器布置方案, 构建较大规模耳周传感默念口令数据集以及提出建模耳周 EMG 信号与默念口令之间时空映射的模型, 进而提出了基于耳周 EMG 信号的默念口令识别方法。本节通过评估模型准确性和稳定性来检验本文方法。具体来说, 首先基于数据集比较本文方法与其他先进且相关方法的性能, 进而评估命令变形、头部运动以及不同时空分辨率与训练数据规模等变量对本文方法的影响。通过以上评

估,旨在展示开发这类系统的可行性及其进一步实际应用的潜力。

3.1.1 默念口令识别

本节将所提出的模型与在类似任务中表现出色的模型进行比较。

1) 考虑到 VGG 在 TensorFlow 语音识别挑战赛中的优异表现及其相对轻量级的架构,本文采用 VGG-16 作为对比模型之一。

2) LSTM 结构普遍用于面部 EMG 信号解码方法中。因此对比模型之一采用文献 [30] 中用于默念语音识别的双向 LSTM 模型 (BLSTM)。

3) 文献 [35] 采用残差网络对 15 种面部微表情进行跟踪,并达到 24.2 mm 的平均绝对跟踪误差。因此,对比模型之一采用该文献所提出的残差网络结构 (Resnet-34)。

4) CNN 在基于 EMG 信号的模式识别任务中达到了目前最佳的性能。因此,对比模型采用文献 [33] 中提出的用于手势识别的多流 CNN 网络 (Multi-view) 以及文献 [29] 中基于肌电信号的面部动作识别 CNN 模型 (BioFace)。

本节采用准确率与 F1 分数 (F1-Score) 作为评价度量。对于对比方法,所有模型的超参数均使用验证集进行调整以达到最优,并采用 Adam 优化器进行 100 epoch 迭代训练。初始学习率设为 0.001。对于训练集与验证集划分,对每名受试者采用 10 折交叉验证并对准确率进行平均作为该对象的对应准确率,最终平均所有对象的准确率作为模型的测试结果。准确率比较结果如图 7 所示。

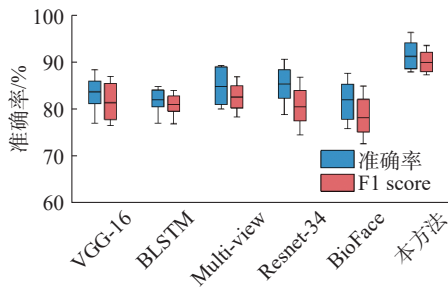


图 7 本方法与对比方法准确率比较
Fig. 7 Accuracy comparison between models

VGG-16、BLSTM、Multi-view、Resnet-34、BioFace 以及本方法的平均精度分别为 83.75%、82.56%、84.84%、85.40%、82.02% 和 91.18%。以上结果得出本方法达到最高的平均分类精度。BLSTM 的结果表明,构建时间建模的 LSTM 网络可能不太适配默念口令识别任务,尤其考虑到相对较低的 SNR 以及测量通道之间的复杂空间映射。VGG-16、Multi-view、Resnet-34 和 BioFace 的性能优于

BLSTM 结构,其可能归因于模型内 CNN 的局部时空特征提取功能。本文架构使用卷积层提取多通道 EMG 的局部时空特征,并基于 Transformer 编码结构提取特征全局上下文时间关系,从而有助于达到更好的性能。此外,尽管 BioFace 与 Resnet-34 在其工作中以较高精度跟踪面部动作,但其性能在本研究任务中相对较差。这可能归因于多种口令近似的口部动作以及耳周传感器布置方案,也表明默念口令识别任务相比 FGR 任务具有更精细的要求。

3.1.2 口令变形泛化

本节测试受试者在口令执行过程中添加语气助词情况下模型的泛化能力,例如模型是否可将变形口令“停止吧”或“停止啊”识别为“停止”。为了验证此性能,在数据采集过程中要求 45 名受试者中的 8 名额外执行一组数据采集,口令执行过程中各以标准或添加语气助词的方式,每种方式执行 10 次,并随机顺序。在模型测试过程中,该组数据作为测试集用以并评估口令变形的影响。

图 8 给出了 8 名受试者在标准和变形口令条件下的平均分类准确率比较,分别为 91.83% 和 88.91%。标准口令条件下的平均准确率与模型在整个数据集上进行测试的结果相似。尽管变形口令条件下的平均准确率略低于标准口令条件,但精度仍处于可接受的范围,这表明本文算法在口令变形时仍具有良好的泛化能力。

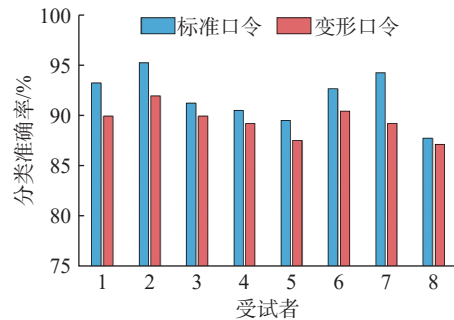


图 8 命令变形下的分类精度比较
Fig. 8 Accuracy comparison under command deformation

3.1.3 头部动作泛化

本节评估模型在头部轻微晃动情况下的泛化性能。为了验证此性能,在数据采集过程中要求 45 名受试者中的 10 名额外执行一组数据采集,并在执行口令过程中随机选择轻微的点头或摇头,每种口令执行 20 次。图 9 给出了 10 名受试者在头部静止和头部晃动条件下的平均准确率比较,分别为 91.21% 和 90.85%。从结果可以得出,头部静止与头部晃动条件下的平均准确率相

似, 证明本文方法在头部晃动情况下的性能相对稳定。

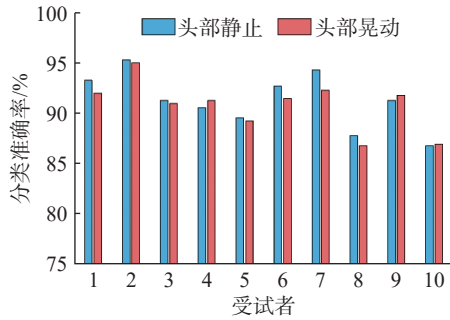


图 9 头部运动下的分类精度比较

Fig. 9 Accuracy comparison under head motions

3.1.4 信号时空分辨率泛化

为系统评估算法对不同传感系统配置的适应性, 本研究通过控制变量法对传感器数量(空间分辨率)和采样频率(时间分辨率)进行多维度性能测试。空间分辨率测试涵盖双侧耳周区域 1 至 3 个传感器的所有组合, 时间分辨率测试选择硬件兼容的 4 种采样率(250、500、750、1 000 Hz)。针对 250 Hz 采样率特有的信号特征, 将滑动窗口延长至 4000 ms 以规避模型末层卷积核的维度冲突问题。此外, 采用单因素方差分析(one-way ANOVA)对性能进行统计验证, 针对不同传感器数量(3 水平)和采样率(4 水平)分别建立独立模型, 并以 $p < 0.05$ 作为显著水平。图 10 给出了在不同时间和空间分辨率下的模型性能, *表示显著差异。

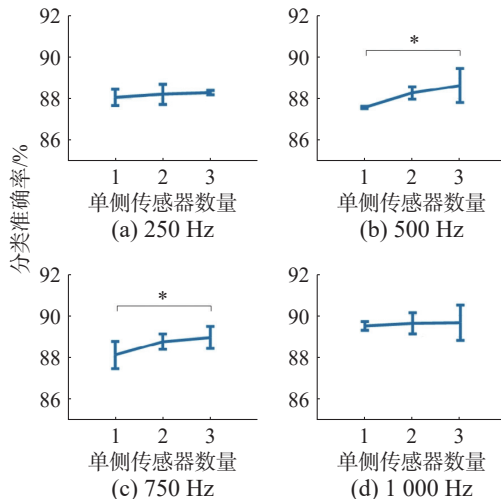


图 10 不同时空分辨率下的精度比较

Fig. 10 Accuracy comparison under different temporal and spatial resolutions

从图 10 可以看出, 除了在 500 Hz 和 750 Hz 采样率配置下使用 1 个和 3 个测量通道之间的准确率差异存在显著性(分别为 $p = 0.03$ 和 $p = 0.04$), 其余时间和空间分辨率下的性能差异均不显著。

这表明不同传感系统的配置差异可能对性能产生轻微影响。在 500 Hz 和 750 Hz 采样率下, 使用 3 个测量通道的准确率显著高于使用 1 个测量通道的准确率, 显示出传感器数量增加对模型性能的提升。上述统计分析表明, 尽管极端配置(如最低空间分辨率+中等采样率)可能产生可检测的统计差异, 但算法在 95% 置信区间内保持相对稳定的性能表现(总体准确率 89%~91%), 验证了方法对传感系统配置变化的强适应性。

3.1.5 训练数据规模

本节评估模型对训练数据规模的敏感性及其任务学习难度, 设计渐进式数据规模实验框架。采用分层抽样法从每位受试者数据中构造 9 种训练集比例(10%~90%, 步长 10%), 各比例训练集均保留原始数据的类别分布特征。为避免单次试验抽样可能带来的偏差, 进行 10 次抽样-训练-测试循环, 并计算平均准确率作为最终结果。单因素方差分析用于分析显著差异($p \leq 0.05$)。图 11 给出了不同训练数据规模下的精度比较, ***表示显著差异。

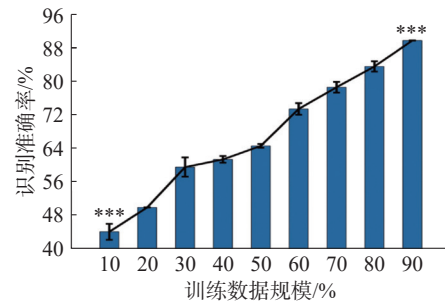


图 11 不同训练数据规模下的精度比较

Fig. 11 Accuracy comparison under different training data scales

如图 11 所示, 随着训练数据量的增加, 准确率也随之提高。每种训练数据规模的平均准确率分别为 44.16%、49.95%、59.55%、61.37%、64.60%、73.39%、78.55%、83.50% 和 91.06%。显著性分析结果显示 90% 训练数据对应结果与其他训练数据规模对应结果具有显著性差异。此外, 从 10% 到 90% 数据规模的性能显著差异表明 EMG 默念口令识别存在显著数据效率瓶颈(达到 80% 以上准确率需至少 80% 训练数据)。此外, 模型性能在低数据域(训练数据规模 < 50%) 呈现强振荡, 反映了 EMG-口令映射存在高度非线性, 这一结果也隐含地证明了默念口令识别任务下模型学习 EMG 与口令之间映射的难度。

3.1.6 人机交互用户接受度调研

本实验最后通过问卷调查的方式评估用户对本文所提出人机交互方式的接受程度。调研对象

为 3.1.2 节至 3.1.5 节所招募的 10 名测试对象, 问卷内容参考人机交互领域广泛应用的用户满意度量表^[35]。该量表根据实验过程中的用户满意度、操作易用性、交互自然性以及用户期望度 4 个方面对本文方法进行评分。例如, 用户满意度衡量受试者在实验过程中对于任务负荷和花费时间的满意度, 交互自然性则评估人机交互提示信息的直观性与易理解性。量表中 7 分为最高分(强烈同意), 1 分为最低分(强烈不同意)。表 4 为针对 10 名用户的统计结果。其中, 每一问题对应分数为该问题下所有受试者的平均得分。从结果可以得出, 用户对于本文默念口令识别任务负荷的满意度、交互系统操作易用性、交互信息的自然性以及方法的期望度较高, 整体用户评分高于 6 分。其中, 任务耗时满意度相对偏低的原因在于本文深度学习需求数据量较高, 因此数据采集时间相对较长。

表 4 人机交互用户满意度调研统计结果

Table 4 Statistical result of user satisfaction survey on human computer interaction

调研内容	问卷问题	用户分数
用户满意度	任务负荷满意度	7.0
	任务耗时满意度	6.2
操作易用性	系统操作负荷度	1.0
	系统操作混乱度	1.0
	系统操作满意度	6.7
交互自然性	交互信息清晰度	7.0
	交互信息易读度	7.0
	交互信息满意度	7.0
用户期望度	系统组件满意度	6.5
	系统功能期望度	6.7

3.2 实验结果分析

本研究旨在提出基于耳周传感的默念口令识别框架设计。与类似的方法和技术相比, 本研究通过考虑社交尴尬最小化和促进微交互因素来限制传感器部署空间。本研究首先确定语音组合原则, 并相应地构建与筛选默念口令。随后, 根据通道信噪比和分类准确率确定最优耳周传感布置方案。最后, 提出 CNN-Transformer 网络构建 EMG 信号与默念口令之间的映射。在评估过程中, 将本文算法与类似任务的先进模型进行比较, 并测试本文方法在非理想条件下如头部运动、指令变形以及不同的空间和时间分辨率的稳定性。本文方法性能提供了耳周传感默念口令识别系统可行性的初步证明。

用户初始化适配 本文基于特定受试者划分

训练集与验证集对比本方法与其他先进方法性能。然而, 开发可适配不同用户信号特征甚至习惯的模型将尤为有益。为训练模型, 本文数据集中每名受试者每个口令类型需重复 40 次, 其缺点在于用户在初次使用该模型时必须执行大量指令用于训练, 这可能会对用户接受度较为不利。能够用少量校准样本调整模型的少样本学习框架将有助于解决这个问题, 本文将其作为未来工作留待研究。

非理想环境泛化 指令变形实验意图模拟日常生活中用户随意执行指令的情况。由于本研究不限制受试者执行默念口令的速度, 因此变形指令的持续时间从 1.2 s 到 2.4 s 不等, 在受试者和训练与测试数据之间也存在差异。因此, 这可能表明本文提出的方法可适应不同的语速。

在日常使用中, 用户配戴耳机进行步行或运动属于常见情况。在这种情况下, 如果传感器相对于皮肤移动, 就会产生运动伪影。与步行相比, 快速的头部运动是一个相对极端的条件, 通常会引起较为严重的运动伪影。在轻微头动实验中的分类结果证明了本方法的稳定性。

为了研究本文方法面向不同可穿戴计算设备的兼容性, 对比本文方法在不同 EMG 空间和时间分辨率下的性能差异以模拟模型针对不同传感系统的泛化情况。即使传感器数量与采样率降为最低, 本方法也可以达到约 87% 的分类准确率。此外, 随着采样率和传感器数量的增加, 模型性能随之提高。因此, 尽管本文方法在稀疏的空间和时间分辨率下表现出较为可观的性能, 但仍建议使用 3 个传感器通道与 1 000 Hz 采样率作为最终配置。

最终, 实验研究本文方法在训练数据获取受限条件下的模型性能。当训练数据量相对较少时(50% 数据规模以下), 本文方法难以达到较为理想的精度。一方面证明了基于耳部肌电的默念口令识别任务难度, 另一方面肌电信号的个体差异性也影响了模型在少量训练样本条件下的性能。针对这一问题, 采用基于半监督或无监督的模型迁移方法可潜在降低数据量需求的同时提升模型的泛化性能。

识别精度提升 本研究限制了传感器布置空间, 但并没有限制用户口型或执行指令的方式。当前研究基于耳周传感实现了约为 90% 的准确率, 足以满足一般使用要求。增加传感器数量以覆盖更多的面部区域将更好地检测到唇动肌肉相关信号, 从而进一步提高性能。此外, 长时间下

的EMG信号的非平稳性问题也将包含在本研究未来工作中。

4 结束语

本文提出基于耳周EMG信号的默念口令识别方法,并设计多流CNN-Transformer网络用于建模EMG信号与默念口令之间的时空映射。通过广泛实验确定构建口令集的经验原则并筛选出最优默念口令与最优传感器布置方案,最终构建包含45名受试者的数据集以训练模型。广泛的实验表明,本文方法以91.18%的平均准确率识别默念口令,优于其他先进模型,并在口令变形和头部晃动情况下表现出较好的稳定性。此外,空间和时间分辨率实验表明本文方法面向不同传感系统的泛化能力。本研究展示了有前景的默念口令识别框架,进而为默念口令识别商业产品的开发奠定基础。

参考文献:

- [1] NICHOLLS B, ANG C S, KANJO E, et al. An EMG-based eating behaviour monitoring system with haptic feedback to promote mindful eating[J]. *Computers in biology and medicine*, 2022, 149: 106068.
- [2] RASHID N, DAUTTA M, TSENG P, et al. HEAR: fog-enabled energy-aware online human eating activity recognition[J]. *IEEE internet of things journal*, 2021, 8(2): 860–868.
- [3] BALLI S, SAĞBAŞ E A, PEKER M. Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm [J]. *Measurement and control*, 2019, 52(1/2): 37–45.
- [4] BI Hongliang, SUN Yuanyuan, LIU Jiajia, et al. Smart-Ear: rhythm-based tap authentication using earphone in information-centric wireless sensor network[J]. *IEEE internet of things journal*, 2022, 9(2): 885–896.
- [5] KHAJAVI M, AHMADYFARD A. Human face aging based on active appearance model using proper feature set[J]. *Signal, image and video processing*, 2023, 17(4): 1465–1473.
- [6] IRAVANTCHI Y, ZHANG Yang, BERNITSAS E, et al. Interferi: gesture sensing using on-body acoustic interferometry[C]//Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Glasgow Scotland: ACM, 2019: 1–13.
- [7] WAN Jun, LI Jing, LAI Zhihui, et al. Robust face alignment by cascaded regression and de-occlusion[J]. *Neural networks*, 2020, 123: 261–272.
- [8] 王贤兵. 人脸识别系统技术分析[J]. *中国科技信息*, 2024(2): 74–76.
- [9] 王海荣. 面部表情追踪技术在VR设备中的发展研究[J]. *电脑知识与技术*, 2019, 15(32): 232–235.
- [10] WANG Hairong. Research on the development of facial expression tracking technology in VR equipment[J]. *Computer knowledge and technology*, 2019, 15(32): 232–235.
- [10] MCMAHAN B, RAO D. Listening to the world improves speech command recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2018, 32(1): 378–385.
- [11] 张锋, 陶浩兵, 慕京生, 等. 基于Bi-LSTM的人机语音交互[J]. *传感器与微系统*, 2022, 41(4): 104–107, 116.
- [11] ZHANG Feng, TAO Haobing, MU Jingsheng, et al. Man-machine voice interaction based on Bi-LSTM[J]. *Transducer and microsystem technologies*, 2022, 41(4): 104–107, 116.
- [12] GAO Yang, JIN Yincheng, LI Jiyang, et al. EchoWhisper[J]. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2020, 4(3): 1–27.
- [13] MU Ruojun, CHEN Jianshe. Oral bio-interfaces: properties and functional roles of salivary multilayer in food oral processing[J]. *Trends in food science & technology*, 2023, 132: 121–131.
- [14] ROSTAMINIA S, LAMSON A, MAJI S, et al. W!NCE: unobtrusive sensing of upper facial action units with EOG-based eyewear[J]. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2019, 3(1): 1–26.
- [15] FAJARDO J M, GOMEZ O, PRIETO F. EMG hand gesture classification using handcrafted and deep features[J]. *Biomedical signal processing and control*, 2021, 63: 102210.
- [16] SIMÃO M, NETO P, GIBARU O. EMG-based online classification of gestures with recurrent neural networks [J]. *Pattern recognition letters*, 2019, 128: 45–51.
- [17] 章毅, 吕嘉仪, 兰星, 等. 结合面部动作单元感知的三维人脸重建算法[J]. *软件学报*, 2024, 35(5): 2176–2191.
- [17] ZHANG Yi, LYU Jiayi, LAN Xing, et al. AU-aware algorithm for 3D facial reconstruction[J]. *Journal of software*, 2024, 35(5): 2176–2191.
- [18] KIM T, SHIN Y, KANG K, et al. Ultrathin crystalline-silicon-based strain gauges with deep learning algorithms for silent speech interfaces[J]. *Nature communications*, 2022, 13(1): 5815.
- [19] 刘万阳, 李晓欧. 基于表面肌电信号的虚拟现实控制系统设计[J]. *计算机测量与控制*, 2019, 27(3): 107–111.
- [19] LIU Wanyang, LI Xiaou. Design of control system for virtual reality based on surface electromyography signal[J]. *Computer measurement & control*, 2019, 27(3): 107–111.

- [20] 孙利, 李放, 王晓梅, 等. 脑电图、肌电图及临床症状学在双侧强直发作局灶性癫痫定侧中的价值评估[J]. 基础医学与临床, 2023, 43(8): 1289–1293.
SUN Li, LI Fang, WANG Xiaomei, et al. Evaluation of EEG, EMG and clinical semiology in the lateral location of bilateral extremities tonic seizures of focal epilepsy[J]. Basic and clinical medicine, 2023, 43(8): 1289–1293.
- [21] 刘声中, 许德章. 一种基于改进 sEMG 活动段检测的手部运动意图识别新方法[J]. 嘉兴学院学报, 2022, 34(6): 105–112.
LIU Shengzhong, XU Dezhong. A new method for hand motion intent recognition based on improved sEMG signal active segment detection[J]. Journal of Jiaxing University, 2022, 34(6): 105–112.
- [22] CHEN Sanyuan, WANG Chengyi, CHEN Zhengyang, et al. WavLM: large-scale self-supervised pre-training for full stack speech processing[J]. *IEEE journal of selected topics in signal processing*, 2022, 16(6): 1505–1518.
- [23] HUEBER T, BENAROYA E L, CHOLLET G, et al. Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips[J]. *Speech communication*, 2010, 52(4): 288–300.
- [24] XU Sijia, YU Jiexiang, GUO Hongshuang, et al. Force-induced ion generation in zwitterionic hydrogels for a sensitive silent-speech sensor[J]. *Nature communications*, 2023, 14(1): 219.
- [25] MELTZNER G S, HEATON J T, DENG Yunbin, et al. Silent speech recognition as an alternative communication device for persons with laryngectomy[J]. *IEEE/ACM transactions on audio, speech, and language processing*, 2017, 25(12): 2386–2398.
- [26] WAND M, JANKE M, SCHULTZ T. Tackling speaking mode varieties in EMG-based speech recognition[J]. *IEEE transactions on bio-medical engineering*, 2014, 61(10): 2515–2526.
- [27] MATTHIES D J C, STRECKER B A, URBAN B. Ear-FieldSensing: a novel in-ear electric field sensing to enrich wearable gesture input through facial expressions [C]//Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Denver: ACM, 2017: 1911–1922.
- [28] NGUYEN A, ALQURASHI R, RAGHEBI Z, et al. In-ear biosignal recording system: a wearable for automatic whole-night sleep staging[C]//Proceedings of the 2016 Workshop on Wearable Systems and Applications. Singapore: ACM, 2016: 19–24.
- [29] WU Yi, KAKARAPARTHI V, LI Zhuohang, et al. BioFace-3D: continuous 3d facial reconstruction through lightweight single-ear biosensors[C]//Proceedings of the 27th Annual International Conference on Mobile Computing and Networking. New Orleans: ACM, 2021: 350–363.
- [30] KIM M, CAO Beiming, MAU T, et al. Speaker-independent silent speech recognition from flesh-point articulatory movements using an LSTM neural network[J]. *IEEE/ACM transactions on audio, speech, and language processing*, 2017, 25(12): 2323–2336.
- [31] ENGLEHART K, HUDGINS B. A robust, real-time control scheme for multifunction myoelectric control[J]. *IEEE transactions on bio-medical engineering*, 2003, 50(7): 848–854.
- [32] DU Yichun, LIN C H, SHYU L Y, et al. Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis[J]. *Expert systems with applications*, 2010, 37(6): 4283–4291.
- [33] WEI Wentao, DAI Qingfeng, WONG Y, et al. Surface-electromyography-based gesture recognition by multi-view deep learning[J]. *IEEE transactions on bio-medical engineering*, 2019, 66(10): 2964–2973.
- [34] MONTAZERIN M, RAHIMIAN E, NADERKHANI F, et al. Transformer-based hand gesture recognition from instantaneous to fused neural decomposition of high-density EMG signals[J]. *Scientific reports*, 2023, 13(1): 11000.
- [35] GAO Yang, WANG Wei, PHOHA V V, et al. EarEcho [J]. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 2019, 3(3): 1–24.

作者简介:



魏柏淳, 助理研究员, 博士, 主要研究方向为人机交互、可穿戴计算与外骨骼机器人。发表学术论文 17 篇。E-mail: bcwei@hit.edu.cn。



姜峰, 教授, 博士, 主要研究方向为视频压缩、可穿戴计算、多智能体博弈、外骨骼机器人。主持国家自然科学基金面上项目 2 项、省部级项目 5 项, 发表学术论文 40 余篇。E-mail: fjiang@hit.edu.cn。



王修来, 教授, 博士, 主要研究方向为人力资源管理与信息不对称、大数据挖掘与分析和数据智能应用。E-mail: wangxiulai@126.com。