

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2023)06-1513-30

论文引用格式: Tao J H, Gong J T, Gao N, Fu S W, Liang S and Yu C. 2023. Human-computer interaction for virtual-real fusion. Journal of Image and Graphics, 28(06): 1513-1542(陶建华, 龚江涛, 高楠, 傅四维, 梁山, 喻纯. 2023. 面向虚实融合的人机交互. 中国图象图形学报, 28(06): 1513-1542)[DOI:10.11834/jig.230020]

面向虚实融合的人机交互

陶建华¹, 龚江涛², 高楠³, 傅四维⁴, 梁山^{5*}, 喻纯³

1. 清华大学自动化系, 北京 100084; 2. 清华大学智能产业研究院, 北京 100084;

3. 清华大学计算机科学与技术系, 北京 100084; 4. 之江实验室, 杭州 311121; 5. 中国科学院自动化研究所, 北京 100190

摘要: 面向虚实融合的人机交互涉及计算机科学、认知心理学、人机工程学、多媒体技术和虚拟现实等领域, 旨在提高人机交互的效率, 同时响应人类认知与情感的需求, 在办公教育、机器人和虚拟/增强现实设备中都有广泛应用。本文从人机交互涉及感知计算、人与机器人交互及协同、个性化人机对话和数据可视化等4个维度系统阐述面向虚实融合人机交互的发展现状。对国内外研究现状进行对比, 展望未来的发展趋势。本文认为兼具可迁移与个性化的感知计算、具备用户行为深度理解的人机协同、用户自适应的对话系统等是本领域的重要研究方向。

关键词: 人机交互(HCI); 感知计算; 人机协同; 对话系统; 数据可视化

Human-computer interaction for virtual-real fusion

Tao Jianhua¹, Gong Jiangtao², Gao Nan³, Fu Siwei⁴, Liang Shan^{5*}, Yu Chun³

1. Department of Automation, Tsinghua University, Beijing 100084, China; 2. Institute for AI Industry Research, Tsinghua University, Beijing 100084, China; 3. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; 4. Zhejiang Laboratory, Hangzhou 311121, China; 5. Institute of Automation, Chinese Academy of Science, Beijing 100190, China

Abstract: Virtual-real human-computer interaction (VR-HCI) is an interdisciplinary field that encompasses human and computer interactions to address human-related cognitive and emotional needs. This interdisciplinary knowledge integrates domains such as computer science, cognitive psychology, ergonomics, multimedia technology, and virtual reality. With the advancement of big data and artificial intelligence, VR-HCI benefits industries like education, healthcare, robotics, and entertainment, and is increasingly recognized as a key supporting technology for metaverse-related development. In recent years, machine learning-based human cognitive and emotional analysis has evolved, particularly in applications like robotics and wearable interaction devices. As a result, VR-HCI has focused on the challenging issue of creating “intelligent” and “anthropomorphic” interaction systems. This literature review examines the growth of VR-HCI from four aspects: perceptual computing, human-machine interaction and coordination, human-computer dialogue interaction, and data visualization. Perceptual computing aims to model human daily life behavior, cognitive processes, and emotional contexts for personalized and efficient human-computer interactions. This discussion covers three perceptual aspects related to pathways, objects, and scenes. Human-machine interaction scenarios involve virtual and real-world integration and perceptual pathways, which are divided into primary perception types: visual-based, sensor-based, and wireless non-contact. Object-based perception is subdivided into personal and group contexts, while scene-based perception is subdivided into

收稿日期: 2023-01-13; 修回日期: 2023-02-23; 预印本日期: 2023-03-02

* 通信作者: 梁山 sliang@nlpr.ia.ac.cn

基金项目: 国家自然科学基金项目(61971419, 62006223, 62002331)

Supported by: National Natural Science Foundation of China (61971419, 62006223, 62002331)

physical behavior and cognitive contexts. Human-machine interaction primarily encompasses technical disciplines such as mechanical and electrical engineering, computer and control science, artificial intelligence, and other related arts or humanistic disciplines like psychology and design. Human-robot interaction can be categorized by functional mechanisms into 1) collaborative operation robots, 2) service and assistance robots, and 3) social, entertainment, and educational robots. Key modules in human-computer dialogue interaction systems include speech recognition, speaker recognition, dialogue system, and speech synthesis. The level of intelligence in these interaction systems can be further enhanced by considering users' inherent characteristics, such as speech pronunciation, preferences, emotions, and other attributes. For human-machine interaction, it mainly involves technical disciplines in relevant to mechanical and electrical engineering, computer and control science, and artificial intelligence, as well as other related arts or humanistic disciplines like psychology and design. Humans-robots interaction can be segmented into three categories in terms of its functional mechanism: 1) collaborative operation robots, 2) service and assistance robots, and 3) social, entertainment and educational robots. For human-computer dialogue interaction, the system consists of such key modules like speech recognition, speaker recognition, dialogue system, and speech synthesis. The microphone sensor can pick up the speech signal, which is then converted to text information through the speech recognition module. The dialogue system can process the text information, understand the user's intention, and generates a reply. Finally, the speech-synthesized module can convert the reply information into speech information, completing the interaction process. In recent years, the level of intelligence of the interaction system can be further improved by combining users' inherent characteristics such as speech pronunciation, preferences, emotions, and other characteristics, optimizing the various modules of the interaction system. For data transformation and visualization, it is benched for performing data cleaning tasks on tabular data, and various tools in R and Python can perform these tasks as well. Many software systems have developed graphical user interfaces to assist users in completing data transformation tasks, such as Microsoft Excel, Tableau Prep Builder, and OpenRefine. Current recommendation-based algorithms interactive systems are beneficial for users transform data easily. Researchers have also developed tools that can transform network structures. We analyze its four aspects of 1) interactive data transformation, 2) data transformation visualization, 3) data table visual comparison, and 4) code visualization in human-computer interaction systems. We identify several future research directions in VR-HCI, namely 1) designing generalized and personalized perceptual computing, 2) building human-machine cooperation with a deep understanding of user behavior, and 3) expanding user-adaptive dialogue systems. For perceptual computing, it still lacks joint perception of multiple devices and individual differences in human behavior perception. Furthermore, most perceptual research can use generalized models, neglecting individual differences, resulting in lower perceptual accuracy, making it difficult to apply in actual settings. Therefore, future perceptual computing research trends are required for multimodal, transferable, personalized, and scalable research. For human-machine interaction and coordination, a systematic approach is necessary for constructing a design for human-machine interaction and collaboration. This approach requires in-depth research on user understanding, construction of interaction datasets, and long-term user experience. For human-computer dialogue interaction, current research mostly focuses on open-domain systems, which use pre-trained models to improve modeling accuracy for emotions, intentions, and knowledge. Future research should be aimed at developing more intelligent human-machine conversations that cater to individual user needs. For data transformation and visualization in HCI, the future directions can be composed of two parts: 1) the intelligence level of data transformation can be improved through interaction for individual data workers on several aspects, e. g. , appropriate algorithms for multiple types of data, recommendations for consistent user behavior and real-time analysis to support massive data. 2) The focus is on the integration of data transformation and visualization among multiple users, including designing collaborative mechanisms, resolving conflicts in data operation, visualizing complex data transformation codes, evaluating the effectiveness of various visualization methods, and recording and displaying multiple human behaviors. In summary, the development of VR-HCI can provide new opportunities and challenges for human-computer interaction towards Metaverse, which has the potential to seamlessly integrate virtual and real worlds.

Key words: human-computer interaction (HCI); perceptual computing; human-machine cooperation; dialogue system; data visualization

0 引言

人机交互技术旨在利用语音、图像和触觉等信息实现人与计算机之间的信息交换,实现虚实空间中的信息传递,是一门与计算机科学、认知心理学、人机工程学、多媒体技术和虚拟现实等密切相关的交叉学科。随着大数据与人工智能技术的发展,人机交互技术在人们的日常生活中得到了广泛应用。近年来,元宇宙概念的快速兴起,面向虚实融合的人机交互是元宇宙系统中的支撑技术之一。

在十三五期间,国家设立多项重大重点项目以支持人机交互方向的研究。例如,基于云计算的移动办公智能交互技术与系统、多模态自然交互的虚实融合开放式实验教学环境以及云端融合的自然交互设备和工具等国家重点研发项目。人机交互相关研究成果对推动新一代互联网、办公教育和医疗康复等领域技术进步都有重要作用。

在各类机器人、可穿戴交互设备中都产生了机器需要深度理解和响应人类认知和情感状态的需求,提升交互系统的智能化与拟人化水平是人机交互领域的研究热点。本文通过综述人机交互在感知计算、人机协同和个性化人机对话等具体领域的最新进展,帮助人们快速了解和熟悉人机交互涉及的多项技术,对面临的机遇和挑战进行梳理,启发相关研究者做出更有价值的研究工作。

本文从人机交互中的感知计算、人与机器人交互及协同、个性化人机对话、人机交互中的数据变换与可视化等维度介绍面向虚实融合的人机交互的研究进展。其中,感知计算从感知路径、感知对象和感知场景3个方面进行展开;人与机器人交互及协同从实际工业应用出发,系统讨论协同操作类机器人、自动驾驶和服务及辅助类机器人3个领域国内外的最新进展及发展趋势;个性化人机对话交互从语音识别、声纹识别、对话系统与语音合成等角度详细讨论对话交互,尤其是个性化交互方面的最新进展;数据变换与可视化重点从交互式数据变换、数据变换可视化、数据表可视对比与代码可视化等方面分析人机交互系统中数据可视化最新的研究进展。

1 国际研究现状

1.1 人机交互中的感知计算

感知计算试图通过对人的日常行为、心理认知与情绪进行建模,以实现个性化的高效交互,是人机交互中的重要研究方向与热点(Yu和Wang, 2020; Zhang等, 2020a; 刘婷婷等, 2021)。面向虚实融合的人机交互场景,根据感知路径,主要分为基于视觉的感知、基于传感器的感知和基于无线非接触式的感知;根据感知对象,主要分为基于个人的感知和基于群体的感知;根据感知场景,主要分为人的物理行为感知和心理认知感知。本节从感知方式、感知对象和感知场景3个方面综述国际上人机交互研究中的感知计算技术。

1.1.1 人机交互中的感知路径

人机交互中的感知路径通常分为3类,即基于视觉的感知、基于传感器的感知和基于无线非接触式的感知。

1) 基于视觉的感知。近年来,大量研究者基于视觉信息来感知和理解人类行为。基于视觉的人类行为感知通常使用视频摄像机等视觉传感设施来捕捉用户行为和環境变化,利用计算机视觉技术(特征提取、图像分割、动作提取和运动跟踪)来分析观察结果以进行模式识别(Yu和Wang, 2020),广泛应用于机器人目标定位、自动驾驶领域。基于视觉的感知建模一般分为图像表征与预测分类两个阶段。Yang和Tian(2017)提出一种超法线向量通用方案,通过将低级多项式聚合为判别表示(可视为Fisher核表征的简化版本),实现了视频序列中的人类活动识别。Jalal等人(2017)提出多融合特征用于线上人类行为识别系统,实现了从连续的深度图序列中识别人类活动。在预测分类阶段,卷积神经网络(convolutional neural network, CNN)是一种常见的深度学习模型,可自动学习部分图像表征,为许多研究者广泛使用(Min等, 2017)。Wang等人(2020a)提出使用嵌入式相机实时感知软体机器人的高分辨率3D形状。摄像头首先捕捉软体内部的视觉模式,通过卷积神经网络CNN生成表示变形状态的潜在模式,并用该模式重建身体3D形状。基于视觉的感知比较直观且包含丰富信息,但在光线较弱情况下感知性能有限(如恶劣天气下自动驾驶车辆难以正确判

断环境),且易产生隐私问题。

2)基于传感器的感知。随着物联网和泛在技术的进步,众多研究者开始利用传感器对人类活动进行感知和理解。一般而言,传感器可通过智能穿戴设备(如手环、眼镜和手机等)附着在被观察人身上,或安装在周围物理环境中(如室内环境监测台)。基于智能可穿戴技术的传感器一般用来收集用户生理数据和运动信息。Sano等人(2015)发现智能手环数据(皮肤电信号、加速度计和体表温度)和手机数据可用来预测学生学业表现的成绩、睡眠质量、焦虑程度和抑郁指数等。Di Lascio等人(2018)和Gao等人(2022a)利用皮肤电信号预测学生在课堂中的投入度。Obuchi等人(2020)通过手机的传感数据和神经影像对105名大学生的大脑功能连接性进行研究,观察到学生手机的使用行为特征与腹内侧前额叶皮层和杏仁核连通性之间具有相关性,并较准确地预测出静息状态下的脑功能连接。然而,智能可穿戴传感器不适合感知较复杂的物理运动及需要和外界进行频繁互动的活动(Yu和Wang,2020),并可能对用户产生潜在的佩戴负担并干扰用户正常行为。

传感器的低成本和低能耗技术促进了物理空间中传感器网络的部署和发展,海量传感器被安置在物理环境中的特定物体上,通过WSN(wireless sensor network)技术和基于连接的通信技术,监测环境信息(如温度、湿度和光照强度传感器)和用户与环境的互动(如门窗状态传感器、动作传感器)。Gao等人(2021)通过室内环境传感器(如温度、湿、空气流速和二氧化碳传感器等)对来自多个城市的用户热舒适度进行建模,并提出一套迁移学习框架以解决传统热舒适度模型标记短缺的问题。Arief-Ang等人(2018)提出一种基于二氧化碳浓度的半监督学习方法,可仅通过二氧化碳传感器采集的小样本数据(如一天的数据信息)精准预估一个房间内的人员数量。

3)基于无线非接触式的感知。无线电磁波信号如WiFi(wireless fidelity)、RFID(radio-frequency identification)、蓝牙在环境中传播时,由于目标对象的存在会产生反射、散射及衍射等现象。而无线接收端的信号由于目标对象的影响,会产生幅度、相位等特征的变化。研究者通过监测和分析这种变化特征,可实现对感知目标行为的非接触式感知。Arshad等人(2017)提出一种基于WiFi数据包的无

线传感器系统,利用WiFi信号的所有可用子载波,并结合它们相位和幅度的变化,可实现精准的人类行为识别。Shao等人(2017)提出了一种无线、隐形的门禁系统,利用蓝牙低功耗信标的RSSI(received signal strength indicator)信号识别进门来客,准确率最高可达95%。Adib和Katabi(2013)发现人类可以在不携带任何传输设备的情况下将消息传送到无线接收器(如图1所示),通过对WiFi信号进行分析,可以识别墙壁后或关门后的移动物体,并识别封闭房间内的人数及其相对位置。

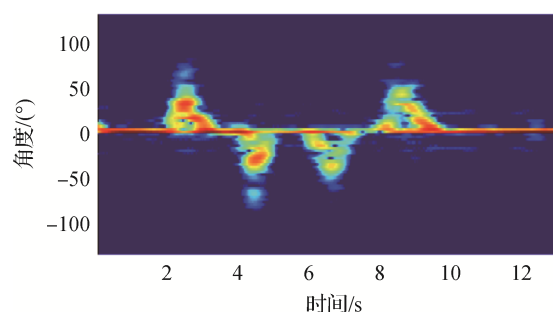


图1 通过WiFi检测到的人体姿势(Adib和Katabi,2013)

Fig. 1 The human gesture detected by WiFi

(Adib and Katabi, 2013)

1.1.2 人机交互中的感知对象

1)基于个人的感知。感知计算领域绝大部分研究工作的感知对象为个人,一些常见的任务场景为基于地理标记的移动轨迹预测、基于智能设备的交互行为感知以及基于无线信号的个人身份识别等(Yu和Wang,2020)。通过感知用户操作或状态,管理者可以调整环境(如调节手机音量、调整房间照明强度)或提供个性化服务(如音乐推荐)。Sadri等人(2018)通过用户历史轨迹数据和当天早晨的初始轨迹,利用马尔可夫模型和长短期记忆神经网络,精准预测出当天晚些时候用户的完整连续轨迹(如未来位置的顺序、停留时间和出发时间)。Huynh等人(2018)使用智能手机(记录触摸事件)、智能手环(记录光电容积描记和皮肤电活动传感器信号)和外部深度相机(记录骨骼运动信息)对手机游戏玩家的投入度水平进行精准的感知预测。

2)基于群体的感知。随着传感器网络在公共设施中的大规模部署和带有地理标记的智能设备的广泛普及,越来越多的研究者开始致力于将“数字足迹”编译为人们日常生活的全面图景,并在多种任务上实现了基于群体的智能感知,如人群情绪识别、区

域安全治理、公共交通规划和城市环境监测等。Bogomolov 等人(2014)提出一种用多模态数据源预测地理空间中犯罪的方法,通过使用移动网络中的匿名人群行为数据来解决犯罪预测问题,并在伦敦真实犯罪数据集中以70%的准确度成功预测犯罪热点。近些年,随着移动互联网的飞速发展,越来越多的人类群体活动迁移至网络平台。Dey 等人(1999)开发了一款会议助理,可通过上下文感知和可穿戴技术,感知到参会者的状态并协助用户间的交流互动。Huang 等人(2022)利用社交媒体 Twitter 平台的数据进行短文本主题识别,以分析美国纽约州公园在新冠疫情发生前后的公共需求变化和价值趋势。

1.1.3 人机交互中的感知场景

人机交互中的感知场景分为两类,即人的物理行为感知和人的心理认知感知。

1)人的物理行为感知。人的物理行为感知与建模是一个亟待解决的问题,其通过传感器、智能设备或视觉的方式采集信息,并在智能家居、智能驾驶、老年人护理和医疗保健等领域得到广泛应用(Jobanputra 等,2019;Shao 等,2018)。研究者通过采集多模态数据,利用机器学习或搭建神经网络的方式,对人类日常活动信息(如步行、跑步、睡觉、驾驶和烹饪等)进行建模和识别。Paul 和 George(2015)使用KNN(K-nearest neighbors)算法和基于聚类的改进算法,通过安卓手机上的加速度计数据对人类日常活动行为进行了准确识别。Ronao 和 Cho(2016)提出一种基于智能手机传感器的人类行为识别深度卷积神经网络,通过导出更相关且复杂的特征,在人类动态活动场景下实现了几乎完美的分类性能。

2)人的心理认知感知。近些年,智能可穿戴和移动设备不断普及,众多研究者致力于研究人们日常生活中的性格特质、心理健康、认知符合和情绪状态。智能设备使用行为数据可用于预测用户性格,以方便开发者服务不同类型用户。Gao 等人(2019)发现使用手机通讯信息和加速度传感器数据可高效预测用户性格特质。多模态传感器数据(如皮肤电信号、血容量脉搏和脑电信号等)也广泛应用于用户的情绪识别和监测。Pakarinen 等人(2019)用皮肤电信号预测用户自我评估的焦虑和心理愉悦程度。Lin 等人(2010)用脑电信号识别人在听音乐过程中的情绪变化情况。Gao 等人(2020b)用皮肤电信号、血容积脉、加速度数据和体表温度等对高中学生的

课堂认知、行为和情感专注度进行了建模预测。Wang 等人(2018)对智能手机使用行为和手环数据进行特征提取,可精准预测大学生日常抑郁情绪变化。

1.2 人与机器人交互及协同

人与机器人的交互(human-robot interaction, HRI)是一个正在高速发展的交叉学科研究领域,主要涉及机械与电气工程、计算机和控制科学以及人工智能等技术学科,也包括丰富的心理学、设计学等人文学科(Singer, 2009; Sheridan, 2016)。最早的机器人实践研究是从军事需求开始的,被誉为自原子弹爆炸以来军队内部最大的革命(Singer, 2009)。随着技术的进步,机器人领域的技术也逐渐转为民用目的,使汽车自动化、家庭自动化和医疗自动化成为可能。近年来,相关领域的文献数量在迅速增长(Singer, 2009)。与人交互的机器人,按功能目的可以分为协同操作类机器人、服务及辅助类机器人、社交娱乐及教育类机器人3类。

1.2.1 协同操作类机器人

在工业4.0背景下,机器正在变得智能化,越来越多的机器人被添加到劳动力队伍中,以提高过程质量和生产率。然而面对复杂的场景和需求,人力仍然是必要的,以积极响应长尾的极端场景和不断变化的市场需求。其中,典型的协同操作类机器人有工业协作机器人、自动驾驶及人机共驾车辆和遥控操作机器人。

1)工业协作机器人。工业机器人主要用来替代或协助人类以高精度执行各种重复/危险和繁琐的制造任务,在过去30年中,这样的工业机器人已经取得了很大进展(Hentout 等,2019)。然而受限于目前的技术能力,完全自主的工业机器人还是困难的,这是因为有些任务可能过于复杂无法完全由机器人完成,或过于昂贵无法完全自动化。人类工作者与机器人共同执行这些任务是最灵活和可负担得起的解决方案,所以协作机器人(collaborative robotic, cobot)用来在人类帮助下执行多种复杂生产任务(Galin 和 Meshcheryakov, 2021)。

为了在真实的工业环境中实现操作员和机器人之间流畅的通信和工作流,需要为不同的工作场景设计反应方法来响应不确定性和意想不到的行为,Rodríguez-Guerra 等人(2021)设计了一个新的4个工业工程级别的分类法,如图2所示。下层(任务设

计)是最接近于直接的人机交互,上层是直接的操作层收集机器人能够处理的所有动作,这些动作是几个任务的关联。处理不同子流程管理的最后一个级别是工业流程级别,其任务依赖于协调若干自主和协

作的子流程,以完成令人满意的生产。该分类回应了5个已确定的挑战,即物理接触管理、对象处理、环境规避、任务调度和管理、任务调度适应性。前3个挑战属于操作级别,而后两个适合工作单元级别。

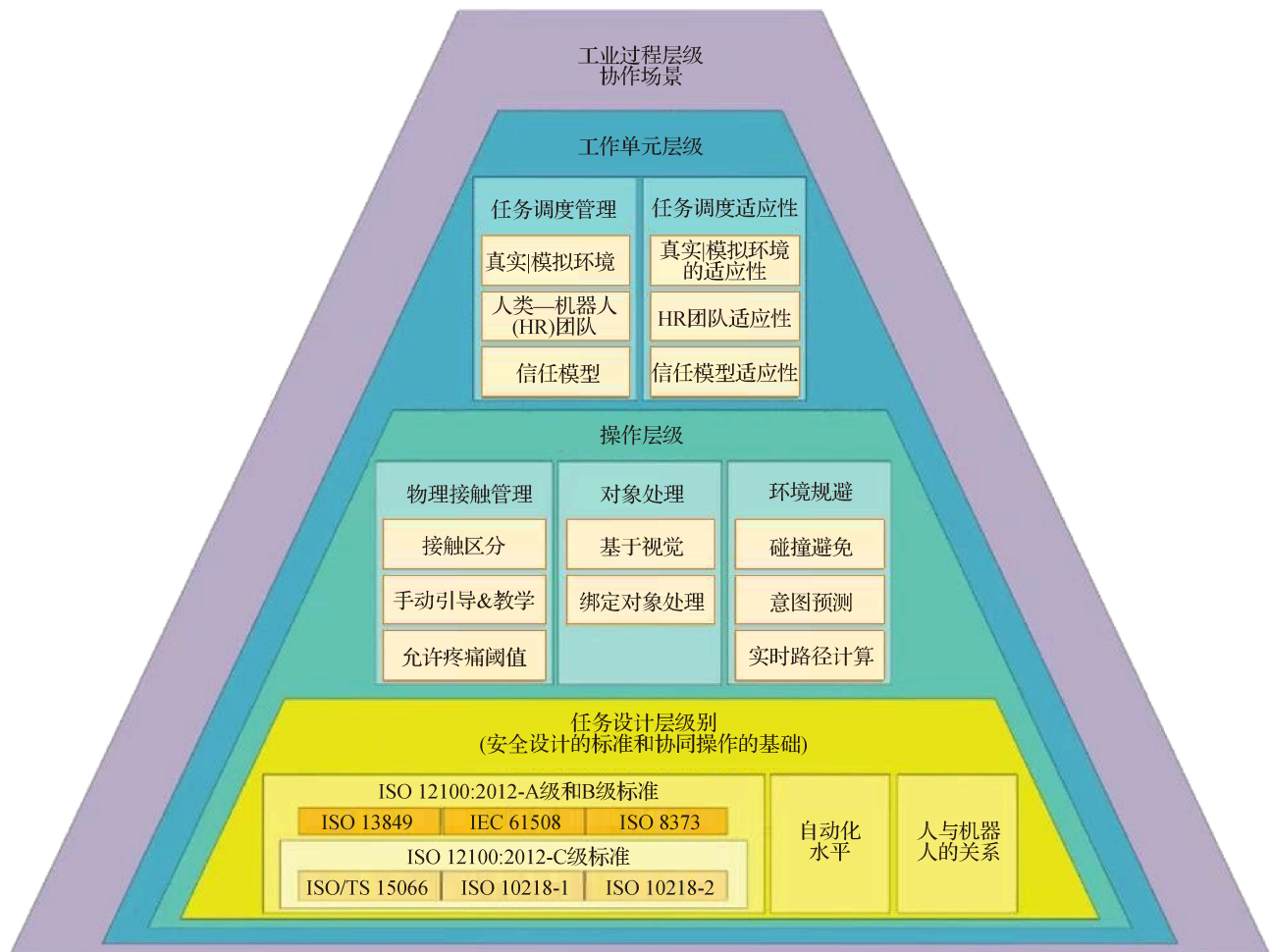


图2 按4个工业过程级别分类的人机协同场景(Rodríguez-Guerra等,2021)

Fig. 2 Human-machine collaboration scenarios classified by four industrial process levels (Rodríguez-Guerra et al., 2021)

2)自动驾驶及人机共驾车辆。自动驾驶与智能汽车人机共驾机器人受到广泛关注。自2020年国务院发布新能源汽车产业发展规划(2021—2035年)以来,我国智能网联汽车产业迎来快速发展期,智能车辆和自动驾驶技术正在经历从硬件、软件、系统工程到基础设施和政策方面的快速发展(萧河,2020)。通过大幅度降低事故率和驾驶过程中人的工作量,自动驾驶技术有望提高车辆安全性,提高交通和环境效率,正在刺激全世界对自动驾驶和驾驶辅助技术的投资(萧河,2020;Casner等,2016)。广泛接受的自动驾驶分级方法之一是由汽车工程学会定义的,该学会将自动驾驶车辆分为6个等级,从L0

级全手动到L5级全自动驾驶(石娟等,2018)。受限于汽车技术各个阶段的发展规律、法律与法规以及道德伦理问题等约束,未来智能汽车将长期处于人机共驾阶段,即车辆的控制权将会在自动驾驶系统和人类驾驶员之间进行切换(L2到L4均存在驾驶权切换)(陈进,2021)。

具有L2—L3级别自动驾驶能力的智能网联汽车与传统汽车相比,驾驶模式已发生显著改变,由单纯的手动驾驶转变为手动驾驶与人机共驾混合存在的新模式,这给社会带来了一系列挑战。美国加州伯克利大学的Wang等人(2020c)提出关于自动驾驶汽车的共享控制有4个显著不同的隐喻。美国麻

省理工学院的 Fridman(2018)从设计和开发以人为中心的自动驾驶系统的角度出发,提出了7大原则,即共享驾驶权、从数据中学习、驾驶人感知、共享感知和控制、深度个性化、不回避设计缺陷和系统级体验。吉林大学的韩嘉懿(2022)针对广义人机共驾描述的宽泛技术架构,按照驾驶人和智能系统的交互关系以及对汽车的操纵方式,将广义人机共驾分为4种形式,分别为后备式、分工式、分时式和融合式。

随着汽车自主性获得越来越多的控制权,驾驶员作为人一车共享控制系统中的重要智能体,需要对其认知过程、控制策略和决策过程进行精确建模。由于人的固有特性,驾驶员与自动驾驶智能体之间的交互策略设计给以人为中心的驾驶辅助系统带来了极大挑战。这产生了许多开放式问题以待更多研究。

3) 遥操作机器人。遥控空间、机载、陆地和海底飞行器,用于在危险或不可接近的环境中执行非常规任务。这样的机器称为遥操作机器人。它们在远程物理环境中执行操作和移动任务,与远程人类的连续控制运动相对应。如果计算机被人类主管间歇性地重新编程以执行整个任务的一部分,那么这样的机器就是遥控机器人。

近年来,远程机器人操作在各个领域得到了广泛应用。在远程外科领域,许多远程机器人操作方法得到了应用(Abbou等,2017)。Hashimoto等人(2011)提出了TouchMe系统,这是一个基于增强现实的远程机器人操作框架。人机界面是一个触摸屏。用户可以通过一个第三人称视角游戏控制移动的机器人和它的机械手。在机器人装配中也采用了远程机器人操作。Wang等人(2014)提出一个远程机器人装配框架,采用远程机器人操作方法实现机器人的实时控制和监测。

美国西北大学提出一种共享自治下辅助遥操作中意图推理的数学公式,递归贝叶斯过滤方法建模并融合多个非语言观察,在没有明确沟通的情况下基于概率推理用户的预期目标(Jain和Argall,2020)。美国乔治亚大学设计并评估了一个沉浸式虚拟现实遥操作系统,提供更自然、更直观的空间控制和远程机器人工作空间的查看,控制双手动机器人机械手进行有效操作(Franzluebbbers和Johnson,2019)。意大利技术研究所针对柔性机械臂的遥阻抗控制问题,受人类运动控制原理的启发,对人类手

臂末端刚度行为进行建模,提出了一种简化的机械臂末端刚度模型(Ajoudani等,2018)。美国英伟达公司开发了一种低成本的基于视觉的遥操作系统DexPilot,只需观察徒手就可以完全控制整个23 DoA(degree-of-actuation)机器人系统(Handa等,2020)。

1.2.2 服务及辅助类机器人

将人工智能整合到服务和辅助机器人中,是迄今为止最令人兴奋且具有颠覆性的做法之一。本节将服务机器人和辅助机器人作为典型类型进行综述。

1) 服务机器人。服务机器人的定义是多样化的,Kuo等人(2017)认为服务机器人是“基于系统的自主和适应性接口,可以与组织的客户进行交互、沟通和提供服务”。定义中包括智能概念,并指出机器人是“物理体现的人工智能代理人,可以采取对物理世界有影响的行动”。Kachouie等人(2014)认为服务机器人主要用于支持独立生活的基本任务,如机动性和导航。随着时间的推移,服务机器人已经纳入了很多日常活动(Bazzano和Lamberti,2018),如酒店管理(Murphy等,2019)和旅游业(Lu等,2019)。

希尔顿酒店与IBM(International Business Machines Corporation)合作开发了人形门房Connie(Ivanov等,2017)。IBM Watson代表了进入认知系统的第一步,并建立在自然语言处理、深度学习和机器学习等几种人工智能技术的组合上,通过每次迭代和交互的结果来帮助改善学习;客房内的酒店陪伴机器人,如日本筑波大学研究了来自Henn-na酒店的Tulie(Osawa等,2017),通过语音识别来响应客人的要求,这种技术允许设备识别语音信息,同时软件理解客人所说的内容并回应客人的要求。在酒店部门使用人工智能技术的另一种形式是开发虚拟代理和/或聊天机器人。例如,澳大利亚南昆士兰大学研究的亚马逊的Echo被瑞典斯德哥尔摩的阿玛兰登号角酒店使用,并充当管家,通过语音识别工作,帮助客人请求客房服务,提供在线信息援助等(Prentice等,2020)。奥地利的老年学研究院对一种自主移动社会服务机器人原型在帮助老年群体时的交互自然性进行探究,通过371天的研究发现,尽管机器效果满足预期,但是机器人更像一个玩具而非功能性机器人(Pripfl等,2016)。美国马凯特大学研究了在临床医生数量不足以满足临床需求的健康中心环境中开发一种负担得起的用于治疗活动的移

动服务机器人(Wilk和Johnson,2014)。

2)辅助机器人。随着机器人在人们日常生活中普遍出现,越来越多的辅助机器人被开发出来帮助解决许多社会挑战性问题,例如残障人士、人口老龄化和医疗保健的需求。包括针对视觉受损人群的导盲机器人(Varela-Aldás等,2020;Xiao等,2021;Wang等,2021b;Guerreiro等,2019)、面向老年人的辅助站坐(Shomin等,2015;李泽新,2022)和辅助行走机器人、面向肢体残疾人的饮食(Bhattacharjee等,2020)和智能轮椅(Zhang等,2022a)机器人等。辅助机器人的形态具有多样化的特点。例如,导盲机器人分为盲杖型机器人(Varela-Aldás等,2020;Slade等,2021)、小车型机器人(Wang等,2021b;Guerreiro等,2019)及四足机器人(Xiao等,2021)。

美国斯坦福大学设计开发了盲杖形态半自动导航机器人,使用横向全向轮配合盲人的推动实现避障和导航(Slade等,2021)。美国加州大学利用美国麻省理工大学的Mini-Cheetah四足机器人,设计了基于伸缩绳索的交互模式的室内导盲。美国卡耐基梅隆大学与IBM在2019年提出了导盲行李箱,利用雷达与深度相机实现了室内导航(Wang等,2021b)。美国卡耐基梅隆大学基于自平衡机器人,研发了对下肢力量不足人群的站坐交互辅助机器人(Shomin等,2015)。美国华盛顿大学西雅图分校面向行动受损人群的自动喂食机器人,研究了残障人士对辅助机器人的自主性水平的感受和需求(Bhattacharjee等,2020)。英国伦敦大学学院基于智能轮椅与人群的交互需求,展开了参与式设计工作坊的研究,研究表明背景环境对于交互形式具有显著影响(Zhang等,2022a)。

1.2.3 社交及教育机器人

如果要将机器人引入人类环境,机器人代理人不仅需要展示解决问题的复杂方法,还需要发展社会参与相关的行为。随着人工智能和工程技术的不断发展,出现了越来越复杂的社交机器人,不仅出现在电影和电视剧中,而且越来越多地出现在现实世界中(Samani等,2013)。本节重点关注社交机器人、教育机器人、远程呈现机器人3种典型类型机器人的人机交互研究。

1)社交机器人。在第四次工业革命中,社会机器人开始从虚构走向现实。随着复杂的人工智能机器人在日常生活中变得越来越普遍,不同领域的研

究人员正在努力解决人类如何感知与这些机器人互动的问题,以及能在多大程度上将智能机器融入人类的社会环境。这些研究综合了社会和认知神经科学、心理学、人工智能和机器人学,展示了跨越生物科学、社会科学和技术的学科研究如何加深人类对机器人代理人在人类社会生活中的潜力和局限性的理解(Cross等,2019)。

如果要将机器人引入人类环境,就需要赋予它们交互能力。这需要在工程和人工智能领域付出巨大的努力(Prescott等,2019),并涵盖诸如人脸和情感识别及动作和意图预测(Gunes等,2019)、语音处理等许多其他领域(Foster,2019)。世界上第一个社交机器人系统Kismet(Breazeal,2000)是美国麻省理工学院人工智能实验室1997年开发设计的,该系统由低层特征提取系统、高层感知系统、注意系统、动机系统、行为系统和运动系统6个子系统组成。NAO机器人(Gouaillier等,2009)是日本软银机器人公司研发生产的一款仿人社交机器人,于2006年上市,由于价格低廉、易于编程,不仅体积小、易便携,还可以在实验室之外的场地开展研究,因而成为世界众多学术机构广泛使用的机器人平台。

澳大利亚墨尔本大学针对面向独居的老年人的陪伴机器人,展开了老年人对机器人的自主感、对生活的控制感、尊严的维护等相关问题的调研(Coghlan等,2021)。丹麦南丹麦大学设计了一种猫外形的扫地机器人与在养老院的老人进行互动,将具有普通任务功能的扫地机器人与额外的社会角色进行整合,使机器人更容易被接受(Marchetti等,2022)。美国威斯康星大学麦迪逊分校探究了远程呈现机器人的高度及操作员相对于本地用户的感知高度如何影响机器人作为媒介的通信(Rae等,2013)。西班牙德里阿夫达大学专门设计了一个新的社会机器人,在日常生活中协助和陪伴老年人,提供安全、娱乐、个人帮助和激励等领域服务(Salichs等,2020)。

2)教育机器人。社交机器人可以用于教育中作为导师或同伴学习者,已被证明在增加认知和情感结果方面是有效的,并且已经取得了类似于人类辅导限制性任务的结果(Gorham,1988;Witt等,2004)。由于积极的学习结果是由机器人的物理存在驱动的,在教育学的研究中使用的机器人范围很广,从小型玩具机器人到全尺寸机器人。机器人在教育中的功效是人们最感兴趣的,物理机器人也更有可能是

用户那里引发有利于学习的社会行为(Kennedy等, 2015)。在合作任务(Kidd和Breazeal, 2004; Wainer等, 2007; Köse等, 2015)中, 机器人可以比虚拟代理更有吸引力和更有趣, 并且通常被认为是更积极的(Wainer等, 2007; Powers等, 2007; Li, 2015)。对于辅导系统来说, 重要的是, 与同一个机器人的视频表示相比, 物理存在的机器人对其请求产生更多的顺应性, 即使这些请求具有挑战性(Bainbridge等, 2011)。

图3显示了国外已发表研究中广泛使用的机器人。图3(a)是教小朋友下棋的iCat机器人(Leite等, 2011), 图3(b)是帮助孩子改善书法的NAO机器人(Hood等, 2015), 图3(c)是在益智游戏中辅导成年人的机器人(Leyzberg等, 2012), 图3(d)是在日本儿童英语课堂上提供动力的胡椒机器人(Tanaka等, 2015)。



图3 广泛使用的教育机器人类型(Belpaeme等, 2018)

Fig. 3 Types of educational robots widely used (Belpaeme et al., 2018) ((a) iCat robot; (b) NAO robot; (c) tutor robot; (d) Pepper robot)

3) 远程呈现机器人。由于2019冠状病毒疾病大流行和相关的社会疏远措施, 亲身活动大幅减少, 以限制病毒的传播, 带来了严重的孤独感和社会隔离。

远程呈现机器人技术(telepresence robot)是一种能够使人以远程的方式实时地在某一位置借助实体的远程呈现机器人出场, 操作者可借助遥操作系统观察和遥控远程呈现机器人, 使其进行位置移动和物理操作, 以与他人进行社交互动并获得亲身参与感(Gorham, 1988)。作为一种远距离互动的方

式, 远程呈现机器人技术的独特之处在于能够赋予远程用户在遥远地点进行机动操作的能力(Gorham, 1988)。这种移动性有效地增加了远程用户的临场感, 从而为任务协作提供重要支持(Gorham, 1988; Rae等, 2015)。许多研究人员在不同的领域和背景下探索了远程呈现机器人的潜力, 如远程会议协作(Khadri, 2021)、远程场馆浏览(Roberts和Arnold, 2012)、人际沟通(Ogawa等, 2011)和远程医疗护理(Daruwalla等, 2010)等。

美国微软雷德蒙德研究院的相关研究验证了远程呈现机器人促进人际沟通增强任务协作的优势(Rae等, 2015)。美国威斯康辛大学和瑞典厄勒布鲁大学的研究也表明远程临场感机器人的肢体性和移动性可以增强休闲交流、协同任务执行(Rae等, 2014)以及当地人对远程人员的意识和注意力(Coradeschi等, 2011)。瑞典厄勒布罗大学的Kristoffersson等人(2013)的研究表明, 与通过带有静态显示的视频会议系统进行交互相比, 远程呈现机器人可触发参与者更高水平的参与度、兴奋程度、注意力和投入程度(Soares等, 2017)。

1.3 个性化人机对话交互

典型的人机对话交互系统通常包括语音识别、声纹识别、对话系统和语音合成等关键模块, 如图4所示。由麦克风传感器拾取语音信号, 通过语音识别模块转化为文本信息。对话系统通过对文本信息进行处理, 理解用户意图并生成回复。最后, 语音合成模块将回复信息转化成语音信息, 完成交互过程。近几年来, 如何结合用户的语音发音特点、偏好和情感等固有特点, 并针对性地优化交互系统的各个模块, 以提升交互系统的智能化水平, 得到了业界的广泛重视。

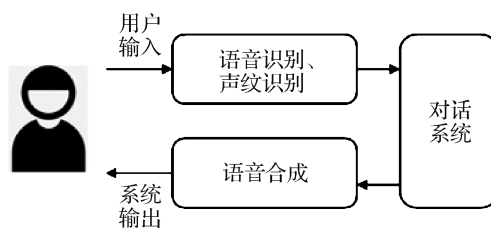


图4 人机对话交互框架图

Fig. 4 The diagram of human-computer dialogue interaction

1.3.1 语音识别

语音是人类最自然、最高效的交互方式, 因此语音识别技术是人机交互的重要入口之一。自20世

纪50年代至今,语音识别一直是人工智能领域的重要研究方向。整体来讲,语音识别技术的发展可以分为3个阶段,包括20世纪70年代以动态时间规整(dynamic time warping, DTW)(Sakoe和Chiba,1978; Müller,2007)和线性预测编码(linear predictive coding, LPC)(Itakura,1975)等技术为代表的模板匹配阶段、20世纪80年代发展的由混合高斯模型—隐马尔可夫模型(Gaussian mixture model-hidden Markov model, GMM-HMM)(Rabiner和Juang,1993;Gales和Young,2008;Gauvain和Lee,1992,1994)主导的统计模型阶段以及2010年前后由神经网络(neural network, NN)引领的深度学习阶段(Mohamed等,2010; Seide等,2011;Graves等,2013)。

随着深度学习技术的发展以及语音数据的逐步完备,语音识别技术取得了巨大进展,开始广泛应用于各种硬件设备。近几年,基于自监督(Zhai等,2019;Kahn等,2020)、半监督(Chung等,2019;Karita等,2018)和无监督学习(Yeh等,2018;Baevski等,2021)方法在语音识别领域受到广泛关注,成为降低语音识别系统对标注数据依赖的有效方法,以实现在许多低资源交换场景下的应用。为了适配更广泛的交互场景,低延迟语音识别是近几年的研究热点之一。流式语音识别与非自回归语音识别是两种系统加速的常用方法。其中,流式语音识别主要包括基于RNN-Transducer(recurrent neural network-transducer)模型的改进框架(Zhang等,2020b;Yeh等,2019;Huang等,2020;Guo等,2021)和基于注意力机制的编解码模型(Inaguma等,2020)。非自回归语音识别模型试图通过编码器预测初始标签,在解码过程中进行纠错或补全(Chi等,2020;Higuchi等,2021);或基于编码器的声学状态,预测完整的输出序列(Chen等,2020a)。

如何根据用户语音的固有特点,对识别模型进行自适应来进一步提高识别的准确率是另外一个研究热点,在智能家居、智慧驾舱和会议设备等诸多终端中都有明确的应用场景。模型微调通过预先训练通用模型,然后根据说话人少量数据对模型整体或局部结构进行微调是说话人自适应的一种较为直接的研究思路(Liao,2013;Yu等,2013)。另一种思路是在模型训练阶段将说话人表示向量(例如i-vector(Dehak等,2011)、d-vector(Wang等,2018)与x-vector(Snyder等,2018)特征)与特征向量拼接,用

以提升识别模型对说话人的适应能力。此外,基于端到端语音识别模型的说话人自适应方法也被提出。例如,不同特征空间的自适应方法(Tomashenko和Estève,2018)、多路径自适应方法(Ochiai等,2018)以及基于注意力的端到端说话人辅助特征建模方法(Delcroix等,2018)等。

1.3.2 声纹识别

由于每个人发声器官存在先天的差异,再加之年龄、性格和言语习惯等各种后天因素的影响,声纹特征可以看做是唯一且在相对长的时间里保持稳定的特征。声纹识别或说话人识别,旨在从语音信号中提取可以表征说话人个性信息的声纹特征或表征,采用模式识别技术自动识别说话人身份。在个性化语音交互中,声纹识别用以识别用户身份信息,实现只与固定用户的交互功能,在智能家居、智慧驾舱和信息安全等领域有越来越多的应用。

在声纹特征的研究方面,代表性的方法包括倒谱技术(Luck,1969)、快速傅里叶变换(Johnsson和Krawitz,1992)、线性预测倒谱系数(linear predictive cepstrum coefficient, LPCC)(Atal,1974)和梅尔频率倒谱系数(Mel-frequency cepstrum coefficient, MFCC)(Davis和Mermelstein,1980)等,其中MFCC应用最为广泛。在方法上,20世纪90年代提出的高斯混合模型(Reynolds和Rose,1995)以及高斯混合模型—通用背景模型(Reynolds等,2000)具有简单灵活和鲁棒性强的优点,是声纹识别通用的方法。21世纪以后,联合因子分析技术(joint factor analysis, JFA)(Kenny等,2007;Dehak等,2011)等算法进一步提升了声纹识别在复杂背景场景下的鲁棒性。

2014年以后,深度神经网络(deep neural networks, DNN)用以提取声纹表征,开始受到关注并得到广泛研究。通常将最后一层隐藏层激活后的输出作为说话人帧级别特征(Variani等,2014),一段语音所有帧级别特征取平均后得到该段语音的句子级特征,称之为d-vector。进一步地,端到端训练方法开始应用于声纹识别(Heigold等,2015;Wan等,2018),整体包括特征提取网络和用于决策打分的判决网络。为了优化声纹识别系统在跨域场景下的性能,基于变分自动编码器(variational autoencoder, VAE)的域对抗训练(domain adversarial training, DAT)方法(Finn等,2017;Tu等,2020)和基于生成对抗网络(generative adversarial network, GAN)的域鲁

棒性训练(Rohdin等,2019;Chen等,2020b)等方法得到了充分研究。

1.3.3 对话系统

对话系统试图实现机器与人的交流达到类人的状态,一直是人们追求的理想人机交互方式,近几年有了迅猛发展。基于管道的对话系统涉及自然语言理解、对话管理和自然语言生成等模块。其中,意图理解对对话系统的智能化水平有着非常重要的影响。面向实际复杂对话场景,目前提出了基于无监督迁移学习的跨领域的意图检测模型(Siddhant等,2019)、基于注意力机制的多意图检测模型(Gangadharaiiah和Narayanaswamy,2019)和基于自适应的Graph-Interactive框架多意图检测模型(Qin等,2020a)等。为了实现对用户情感状态进行追踪,以产生更有个性化关怀的对话序列,在Liu等人(2019)的工作中,情感状态被量化编码并输入模型,用以辅助意图理解和生成应答。

端到端的对话系统可以克服传统级联式对话系统的误差传递问题,受到了广泛关注。典型系统包括Madotto等人(2018)提出的基于指针网络的Mem2Seq(memory-to-sequence)模型,Wu等人(2019a)融合骨架循环网络的GLMP(global-to-local memory pointer)模型等。此外,结合预训练模型的对话系统也受到广泛关注。例如,基于BERT(bidirectional encoder representation from transformers)预训练语言模型提升口语理解能力,并将BERT中的知识蒸馏到意图分类模型中,以实现更准确的人机对话(Kim等,2021a;Lai等,2021;Kim等,2021b)。

1.3.4 语音合成

随着深度学习在语音识别领域取得了突破性的进展,基于深度神经网络的语音合成成为主流。代表性的系统包括由谷歌Deepmind研究团队提出的基于深度学习的WavetNet语音生成模型(van den Oord等,2016)和端到端的语音合成系统(Sotelo等,2017;Wang等,2017b)。

为了保证合成语音具备丰富的情感信息,提高对话过程的拟人化水平,目前语音合成中的情感编码方法包括基于独热向量情感表示的cascadeTTS(cascade text-to-speech)系统(An等,2017;Wu和King,2016)和基于全局风格令牌(global style tokens,GST)的TTS模型(Wu等,2019b;Lorenzo-Trueba等,2018)等。

1.4 人机交互中的数据变换与可视化

1.4.1 交互式数据变换

对表格数据执行各种数据变换操作是完成数据清洗任务的基础。许多基于R语言和Python语言的工具库能够执行各类数据变换操作。虽然基于编程语言的工具库功能强大,却对使用者提出了更高的要求,需要熟练使用编程语言且熟悉工具库函数参数。为了降低数据变换的门槛,许多软件系统基于图形化界面辅助用户完成各类数据变换任务,例如Microsoft Excel,Tableau Prep Builder,OpenRefine等。近年来,研究者设计开发了基于推荐算法的交互式系统(Kandel等,2011;Guo等,2011;Drosos等,2020;Jin等,2017;Bigelow等,2019;Abedjan等,2016;Inala和Singh,2017)。例如,Wrangler系统(Kandel等,2011;Guo等,2011)基于用户对行、列及单元格的操作推荐相应的数据变换操作。基于样例编程的范式,Wrex(Drosos等,2020)和Foofah(Jin等,2017)能帮助用户快速完成数据变换。系统会根据用户在界面的操作,智能生成数据变换代码,帮助用户批量处理复杂数据。此外,还有研究者设计了能够对网络结构进行数据变换的工具,如Origraph(Bigelow等,2019)等。另外,有一部分研究如Dataxformer(Abedjan等,2016)和WebRelate(Inala和Singh,2017)等支持从网站获取数据并执行数据变换任务,提升了数据变换中数据获取阶段的效率,这些工作使数据工作者完成数据变换任务更加方便。

1.4.2 数据变换可视化

数据工作者常常需要理解清洗脚本中所执行的具体数据清洗过程,以了解数据是如何发生变化的。例如,在程序复用中,需要学习其他脚本中数据清洗的思路,以修改并应用于自己的数据上;在双重检验中,需要验证其他人执行数据清洗的过程,以确保清洗结果的准确性;在代码维护中,需要整理没有良好文档记录的代码,以避免维护任务中的错误。传统的方法采用纯文本描述数据转换操作的语义,以帮助数据工作者理解数据表格是如何变化的,如在数据清洗的代码中插入解释数据转换操作的注释,用于数据清洗的包/库(如Python中的Pandas包,R中的tidyr、dplyr包等),其官方文档也采用文本的形式解释各转换操作函数的含义与用法。然而,纯文本的描述方式难以帮助数据工作者直观高效地理解数

据转换操作的语义。此外,由于文本过于灵活,描述风格存在多样性,如果文本内容比较含糊,描述不够准确,可能还会导致数据工作者对该转换操作有错误的理解。

一些工作以文本的形式描述数据变换的语义。WrangleDoc(Yang等,2021)通过比较数据变换前、后的数据表差异,帮助用户理解数据变换的语义。Unravel(Shrestha等,2021)使用一段文本为链式结构代码块的每一行作注解,并使用拖拽交互帮助用户纠正链式结构中的错误。基于代码解析结果,一些工作使用新颖的方式展示数据变换代码的语义。例如, Datamations(Pu等,2021)和 Data Tweening(Khan等,2017)通过精心设计的动画来展示复杂数据变换的语义。然而,以上这些工作局限于可视化单步数据变换的语义,并没有针对多步数据变换做相应的优化。虽然多步数据变换可视为多个单步变换的组合,然而用户需要逐一理解每一步细节,效率较低。

1.4.3 数据表可视对比

2维数据表格是一种组织整理数据的有效手段(Niederer等,2018),由于原始表格常常包含“脏”数据,或数据格式、内容等不符合预期目标,因此必须对表格进行数据清洗。在数据清洗过程中,常常需要比对数据表格的变化,以确认是否成功执行了指定的数据转换操作,或根据当前数据表格的变化确定应该执行何种数据转换操作。然而,由于数据表格包含的行、列数据量可能过大,并且数据转换操作导致数据表格发生变化的种类繁多,使得难以纯手工地对比数据表格在数据清洗前后的变化。因此,一些可视化的工作致力于表格数据可视对比(Furmanova等,2020)。Furmanova等人(2020)将表格数据可视化技术分为3类。1)基于概览技术的可视化(Claessen和van Wijk,2011;Fua等,1999;Yalçın等,2018;Luo等,2019;Wei等,2022c),侧重于展示表格中的数据内容。2)基于投影技术的可视化(Liu等,2017),主要针对高维数据。3)基于表格隐喻的可视化(Gratzl等,2013;Pajer等,2017),在原有表格展示的基础上,可视化数据内容和类型。由于不同技术各有优劣,有些工作采取多种可视化类型(Furmanova等,2020;Lex等,2011,2012;Stahnke等,2016)来提升可视化表达的效果。以上的工作着重于对数据变换操作的可视化,缺乏表格内容变化的

展示。TACO(Niederer等,2018)通过使用时间轴概览、差异直方图和热力图等技术可视化两张表格之间的差异,但是这些技术并不适用于可视化数据变换任务中的语义。

1.4.4 代码可视化

目前有不少程序可视化研究工作旨在可视化实际程序代码的执行逻辑或数据结构(Price等,1993),这些工作按照设计目的可分为代码调试(Kosower等,2014;Cheon等,2015;Moseler等,2022;Kumar等,2021;Jbara等,2019;Hori等,2019)和学习教育(Guo,2013;Khaloo等,2017;Hansen等,2002;Demetrescu等,2002;Balogh和Beszédés,2013)两大类。部分可视化工作旨在帮助用户调试代码。例如,流程图自动生成工具(Kosower等,2014;Cheon等,2015)能够展示代码的执行逻辑,从而帮助用户调试代码。Moseler等人(2022)专注于可视化多线程迸发的程序,通过可视化各个线程中的执行情况,协助用户理解并发式的代码,辅助调试及代码优化。一些工作致力于展示代码中的数据结构(Kumar等,2021)和运行时状态(Jbara等,2019;Hori等,2019)帮助用户调试代码。还有一些可视化工具旨在帮助用户学习编程语言或代码库。Online Python Tutor(Guo,2013)通过可视化代码运行时的状态信息和数据结构,帮助编程新手理解Python程序运行的过程。Khaloo等人(2017)使用VR(virtual reality)技术可视化代码库,相比于阅读传统文档,沉浸式可视化更具有吸引力。有的可视化工具(Hansen等,2002;Demetrescu等,2002)通过展现每步代码的行为帮助用户学习算法。CodeMetropolis(Balogh和Beszédés,2013)则是将代码中的类、变量等信息可视化为3维实体,更为形象地展示代码的层次结构及其属性。

2 国内研究进展

2.1 人机交互中的感知计算

2.1.1 人机交互中的感知路径

国内对于基于视觉、传感器和无线非接触式感知的研究较多,然而对基于传感器和无线非接触感知的研究大多集中在实验室环境而非真实物理环境。在视觉感知领域,厦门大学智能多媒体实验室的Zhong等人(2017)提出一种自动驾驶中目标检测

的特定类别目标排序方法,首先提取特征,包括语义分割、立体信息、上下文信息、基于CNN的对象性和低级提示等,然后使用结构化支持向量机(support vector machine, SVM)学习的特定类权重对它们进行评分。基于传感器的感知领域,清华大学普适计算实验室与人机交互实验室的Liang等人(2021)提出一种环形输入设备,可以通过惯性测量单元(inertial measurement unit, IMU)传感器精确捕捉用户手指运动和手势信息,能够感知相对于地面的绝对手势以及手部的相对运动。在无线非接触式的感知领域,北京大学的张大庆团队在呼吸检测、跌倒检测、睡眠监测、入侵检测、手势识别和室内定位等方面开展了一系列工作。Zhang等人(2020a)首次将低功耗和远距离通信的LoRa(short for long range)技术应用于远距离非接触式感知,利用LoRa接收端的双天线和独特的降噪算法减少信号不同步的误差,并大幅提升感知范围,可在15 m(隔两堵墙)或25 m的情况下检测人体呼吸。Niu等人(2018)首次将菲涅耳区衍射模型应用于人体呼吸检测,并解决了定量计算第一菲涅耳区无线传感信号与人体活动的关系的难题。浙江大学的Yang等人(2014)探索了单户住宅中的运动传感器场景和大学实验室中的电表信息,发现服务商提供的传感信息有助于对居住率甚至当前居住者的身份进行感知预测,且效果明显高于传统的非传感器预测方法。

2.1.2 人机交互中的感知对象

国内对于人机交互中的不同感知对象(个人和群体)研究较多,与国际前沿相差不大。西北工业大学普适与智能计算研究所的Guo等人(2016)提出一种基于群体的群智感知系统,该系统考虑了群体活动的复杂性和动态性,并介绍了一种帮助团体活动准备的智能策略,包括用于宣传公共活动的基于启发式规则的机制和用于私人团体的基于上下文的方法。微软亚洲研究院Zheng等人(2008)提出一种基于监督学习的方法,可从人们的全球定位系统(global positioning system, GPS)日志中推断人们的运动模式,并提出一种图像后期处理算法,以概率方式考虑了现实世界常识约束和基于位置的典型用户行为,可进一步提高推理性能。Wang等人(2017a)提出一种基于地理位置(GPS数据)的移动目的地预测方法,该方法只研究查询轨迹本身的行为,并不匹配历史轨迹,可用于稀疏数据集的目的地预测,具有

较高的有效性和可拓展性。

2.1.3 人机交互中的感知路径

国内在人的物理行为感知领域进展较为迅速,与国际水平相差不大,但多聚焦于算法层面的创新。然而,国内在人的心理感知领域研究相对较少,如感知用户心理疾病、幸福度、工作效率、投入度和体验感等。中国科学院计算技术研究所Lu等人(2022)提出一种基于可泛化传感器的跨域人类活动识别模型,引入了考虑活动语义范围的语义感知方法,以克服域差异带来的语义不一致问题。Qi等人(2018)探索了物联网环境中医疗健康领域的身体活动和检测技术。南京大学Zhang等人(2018)提出了一种基于智能手机感知的复合情绪(多维情绪)机器学习算法,通过光照、加速度计、无线、麦克风、地理位置数据和手机应用使用情况对复合情绪进行预测,平均精确率可达76%。

2.2 人与机器人交互及协同

2.2.1 协同操作类机器人

腾讯机器人X实验室与德国慕尼黑大学、英国利兹大学合作,面向人类与机器人的强力合作过程,规划机器人的配置和抓握动作,对一个物体进行穿刺或切割任务,设计了一种机器人配置,以定位联合操纵的物体,使人类在同一物体上操作的肌肉力量最小(Figueroa等,2021)。山东科技大学基于6自由度协作机器人,提出了一种新的机器人碰撞检测方法,在工业协作机器人和人需要共享工作空间时,保护协同工作时人的安全(胡钰,2020)。武汉理工大学提出面向人机协作安全保障的工业机器人路径规划方法,可以实现工业机器人整体的避障,保障人员安全(李娜,2018)。

吉林大学面向自动驾驶人机交互安全的重要问题,基于L3级别自动车道保持系统,进行了人为误用安全分析与评价和接管策略设计,对提高自动驾驶汽车人机交互安全和接管设计给予参考建议(马海涛,2022)。清华大学Chu等人(2023)面向自动驾驶安全员展开质性研究,对于高风险交通场景人与人工智能的协作和人工智能研发中的伦理问题展开讨论。中西南科技大学的研究报告分析了影响接管决策的因素,基于多资源理论梳理了驾驶任务产生心理负荷的信息加工过程,结合以情境为中心的设计理念与用户体验层级,明确了多模态交接受管提示方式与接管视觉界面的设计目标和方案,为自动

驾驶接管设计提供参考(徐韬,2022)。天津师范大学基于虚拟场景和驾驶模拟器探索了L2自动驾驶情境下驾驶员的心理行为特征,从而为解决接管问题提供一定的依据(徐杨,2021)。

清华大学设计并实现了一个带有新型数据手套YoBu的机械手远程操作系统,可以同时从手臂和手获取人类的运动(Fang等,2015)。广西大学基于主从遥操作机器人展开了运动控制策略及模型优化研究,以支持主从遥操作机器人与新一代信息技术、传统医疗器械深度结合,实现具有超高精度的远程医疗(杨启业,2022)。广东工业大学面向高空、带电和危险电网作业展开了半自主遥操作机器人系统研究,设计开发了一套半自主遥操作电网作业机器人系统,在此基础上,重点解决了针对电网作业中的拧螺母与插开口销两项任务所涉及的技术问题(韦海彬,2022)。中国矿业大学面向微创外科手术机器人,展开辅助柔性针穿刺的遥操作控制方法研究,拓展医生的手术能力(陈肖利,2022)。

2.2.2 服务及辅助类机器人

扬州大学以用户体验设计理论和用户研究方法为理论基础,构建了家庭助老服务机器人产品交互设计框架,为智慧养老提供了一种系统性的新方案,对智能型、服务型机器人的外观设计和互联网智慧居家养老平台的研发提供有益的参考(庞广风,2022)。山东建筑大学运用情境模型理论,对智能服务机器人在场馆情境中的设计方法展开研究,以提升智能服务机器人的可用性及其用户满意度(刘双,2022)。中铁电气化局集团以服务机器人为载体,从机器人与铁路客运站人脸核验、通道门开关、直梯呼叫以及闸机验检票等系统设施融合应用方面展开研究,提出物联方案,通过服务机器人与铁路客运站既有设施物联,实现旅客服务信息的集成共享,扩展机器人服务能力(戴彦华等,2022)。华南理工大学提出了一种移动服务机器人的导航策略,同时整合对人的检测,对机器人实时定位和运动规划,实现有意识、安全、准确、鲁棒和高效的导航(Yuan等,2021)。

百度研究所面向视觉受损人群的导航机器人,提出了基于硬连接的人机运动学模型与基于模型预测的规划和控制算法,以避免盲人与机器或环境的碰撞(Wang等,2021b)。清华大学针对视觉受损人群的自主性需求设计了具有不同自主性的辅助机器人,验证了盲人群体对于出行的控制感的需求

(Zhang等,2023)。上海理工大学等单位结合康复训练与生活辅助机械臂设计需求与特征,提高机械臂工作空间及其运动性能,提出了一种9-DOF(degrees of freedom)上肢康复训练与生活辅助机器人,可用于上肢功能障碍患者的康复训练和日常生活辅助(焦宗琪等,2022)。新疆医科大学第一附属医院研究测试了MAKO机器人辅助下全膝关节置换术治疗伴有严重内翻畸形膝骨关节炎的疗效及可行性,研究表明机器人辅助膝关节置换可以提供个性化手术方案,术中实时反馈提供客观判断,治疗严重内翻畸形的膝骨关节炎可以获得令人满意的短期结果(穆文博等,2022)。沈阳工业大学面向下肢残障或下肢肌肉力量不足的人群站—坐交互中提供辅助的护理机器人,提出了一个由扭转、垂直和水平3个维度构成的具有9自由度的参数模型,旨在模拟站坐交互中人体与座椅接触后的过程(李泽新,2022)。

2.2.3 社交及教育机器人

浙江工业大学利用眼动信号捕捉及分析,将社交机器人应用于孤独症儿童共同注意力的干预(王蒙蒙,2019)。南京师范大学以社交机器人参与网络社交引发的伦理失范现象作为研究对象,采用文献研究法和案例分析法,依托国内外关于社交机器人的研究成果,借鉴新闻传播学、计算机科学、社会心理学和哲学等学科观点,对该现象进行了全面的梳理和分析。

我国针对机器人教育的研究和发展,最初是国家863计划支持下的机器人技术研究和实践(高博俊等,2020)。目前国内已经有多家公司研发出了各式各样的机器人产品,其中具有可视化编程的面向中小学教育的机器人产品中,比较有代表性的包括能力风暴机器人(<http://www.abilix.com/>),优必选阿尔法机器人(<https://m.ubtrobot.com/cn/products/e-bot/>),大疆教育机器人Robomaster S1(<https://www.dji.com/cn/robomaster-s1>)等。

清华大学Gao等人(2022c)结合定制外观的积木智能制造技术,研发物理孪生教育机器人,对学生的学习和情感体验产生了显著影响。合肥工业大学结合机器人操作系统(robot operating system, ROS)和Ubuntu,搭建了一款低成本、高性能和开源的远程呈现移动机器人平台,可实现远程呈现和控制(张华健和钱钧,2019)。北京理工大学提出一种

基于立体视觉与手势识别的远程呈现交互方法,相较于传统方法更灵活和自然,而且可以给用户带来沉浸式的体验(李佩霖,2018)。

2.3 个性化人机对话交互

2.3.1 语音识别

国内在语音识别领域开展了许多杰出的工作。中国科学院自动化研究所提出了基于 speech-transformer 模型的说话人自适应方案(Fan等,2019,2021a)。该方案可以应用到会议场景,提升说话人频繁切换时段的语音识别性能。上海交通大学提出了低资源小语种语音识别以及基于端到端快速口音自适应语音识别方法(Qian和Zhou,2022;Qian等,2022),以适应实际应用场景下的口语自适应问题。中国科学技术大学与科大讯飞联合提出了多个语音识别算法,例如基于改进的推敲网络的端到端语码转换(code-switching,CS)自动语音识别方法(Huang等,2019)。针对实际对话交互场景,西北工业大学与腾讯合作提出了基于对话特征建模以及跨模态特征的语音识别方法(Wei等,2022a,b)。这些进展表明,国内在语音识别领域开始关注实际对话场景及各种场景自适应问题。

2.3.2 声纹识别

清华大学人机语音交互实验室提出了一种简单高效的声纹识别 MFA-Coformer (multi-scale feature aggregation conformer) 结构(Zhang等,2022b)。中国科学技术大学与科大讯飞联合提出的基于深度嵌入学习文本无关说话人识别方法(Li等,2022a),通过联合学习获得对语序不敏感的说话人表征向量;基于类内与类间距离分布对齐策略的方法(Hu等,2022),提高了声纹识别在训练与实际使用环境不匹配情况下的泛化性。此外,中国科学院自动化研究所提出了基于预训练模型的声纹特征提取算法以及应用于会议场景的说话人切换估计(Fan等,2021b,2022)。

2.3.3 对话系统

针对对话过程存在的跨领域问题,哈尔滨工业大学提出了多领域端到端对话系统(Qin等,2020b)。北京大学针对小语料开放域对话系统,提出了基于元学习的训练框架(Song等,2020)。上海交通大学与思必驰公司联合开展了多项工作(Lan等,2018;Chen等,2018),用以提升口语理解、基于强化学习的对话管理性能。清华大学针对开放域对话系统在语言生成、训练数据增广、对话策略等领域

提出了一系列方法,在对话生成中考虑了情感模型、知识表征等(Zhang等,2020c;Shao等,2022;Gao等,2022b),提高对话系统的“拟人化”程度。

2.3.4 语音合成

在语音合成领域,国内诸多研究机构也走到国际前列,尤其在小样本、个性化语言合成领域开展了多项工作。清华大学与字节跳动公司联合提出了合成语音风格建模方法(Li等,2022b)。中国科学技术大学提出了词级语音风格建模方法(Liu等,2022)及一种解耦发音和韵律建模的方法(Peng和Ling,2022),以提高基于元学习的多语言语音合成的性能。中国科学院自动化研究所提出了将目标说话人韵律与音色解耦的合成算法,提高了模型的泛化性(Wang等,2020b)。西北工业大学与腾讯联合提出的零资源语音合成模型 Glow-WaveGan2 (Lei等,2022)中,采用基于变分自编码器的说话人编码器,并构建了连续的说话人空间,可为新说话人生成高质量的语音信号。国内科大讯飞、思必驰、云知声以及百度等企业都推出了个性化语音合成服务,推动了本领域的落地应用和快速发展。

2.4 人机交互中的数据变换与可视化

在数据变换可视化、数据表可视对比及代码可视化方面,国内研究尚在起步阶段。Xiong等人(2022)提出 SOMNUS,通过基于图形图符的节点链接图,来可视化表达数据变换过程。此外,少数工作利用可视化的手段辅助恶意代码检测(龙墨澜和康海燕,2022)。在交互式数据变换方面,国内有不少 BI (business intelligence) 工具,例如阿里的 Quick BI,帆软推出的 FineBI 等均支持交互式数据变换。然而,这些工具需要大量的人工操作,智能化方面较为薄弱。Chen等人(2023)提出 Rigel,使用一种新型的声明式映射方法,减少了数据工作者在完成数据变换操作时与系统交互的次数。Li等人(2023)提出 HiTailor,致力于帮助数据工作者完成多级数据表的变换操作,但这类工作无法处理步骤繁多或操作复杂的数据变换。此外,在用户完成数据转换的过程中仍需要进行交互操作,相关的智能化推荐相对较少。在数据表可视化方面,马楠和袁晓如(2020)通过对数据表的可视化,结合上下文生成相关可视化推荐帮助用户完成可视分析任务。在代码可视化方面,张秀深(2013)提出了一个可视化编程移动学习平台,通过对代码的图形化,辅助用户学习、理解

代码。赵永刚等人(2010)对多线程的代码运行情况进行可视化,帮助用户对代码进行评估。胡珊(2021)将代码可视化应用在高校学生教学工作中,以帮助学生课程学习中的难点。

3 国内外研究进展比较

3.1 人机交互中的感知计算

感知计算技术在国内外外的研究进展如表1所示。总体而言,国内外在感知技术领域的研究进展相似,但研究重点略有不同。在感知路径方面,国内更聚焦于基于视觉的感知计算和基于无线非接触式的感知计算,并在视觉领域已有较成熟的应用场景;国外在基于传感器的感知计算领域发展较为迅速,

且有较多种类的消费级智能可穿戴设备。在感知对象方面,国内外研究发展持平。但是在感知场景方面,国内外研究重点略有不同。国内研究侧重于使用公开数据集进行算法层面的创新,且应用场景颇为单一;国外研究者通常侧重于新颖场景的探索和新问题的定义,通常会在真实世界中收集被试数据,并对被试行为进行感知,学科交叉型更强。同时,国外在心理感知领域的发展较为迅速且呈多样化趋势(如检测多种精神疾病、用户幸福度和工作效率等),而国内在心理感知领域的研究很少,且研究方向较为单一,多注重于传统情绪检测。考虑到感知计算的学科交叉性,国内学者应多与跨学科专家和国外学者进行合作交流,利用国内的技术优势,推动感知计算领域的共同发展。

表1 感知计算技术国内外研究进展对比

Table 1 Comparison of domestic and foreign research progress on sensing technologies

	国内	国外
感知路径	1)基于视觉的感知和基于无线非接触的感知技术和国外水平持平。在基于传感器的感知技术方面有待增强。2)学术界以北京大学为代表,主要研究基于无线非接触式的感知技术;企业界以商汤科技、旷视为代表,主要聚焦于基于视觉的感知。	基于视觉、传感器和无线非接触式的感知技术研究比较成熟,有较多种类的消费级智能可穿戴设备。
感知对象	以清华大学、西北工业大学以及中国科学院自动化研究所为代表,在基于个人和群体的感知层面均达到国际前沿水平。	在个人和群体感知层面发展较成熟,学科交叉性较强。
感知场景	1)研究更偏向于算法技术层面的创新,多采用已知公开数据集;2)在心理感知领域研究较少;3)在物理行为感知领域研究和国外持平。	1)研究更偏向于应用场景的创新;2)研究数据通常来自于真实环境的实地采集;3)近些年来在心理感知领域发展迅速,注重多维度用户心理体验等。

3.2 人与机器人交互及协同

国内工业协作机器人的人机交互研究相对集中于避障和人员安全研究,但是国外的工业协作机器人更加集中于增强人的交互体验,特别是通过触觉传感/反馈的技术,也更强调机器人对于人类协作伙伴的适应性。国内自动驾驶的人机协作研究主要集中于驾驶员的接管问题及相关人因及设计研究,自动驾驶人机协作研究更加多样化,涵盖大规模真实驾驶数据集、自动驾驶与驾驶员分层协作、自动驾驶和行人交互等多个方面。国内遥操作机器人重点面向医疗、电网操作等领域,展开半自主性、辅助遥操作的控制算法研究,国外的相关研究更多偏向于对操作者建模,提供类人的、舒适的和沉浸的遥操作方法。

国内服务机器人的研究更加集中于前期系统设计和实现,国外的相关研究针对于实际部署体

验,甚至长时间的使用评价则更多一些。在辅助机器人研究领域,国内相关研究集中于技术和系统研发,而国外相关研究除了相关技术研发,更进一步地关注到被辅助的用户对自主性、交互体验的需要。

近期国内关于社交机器人相关的研究主要集中于增强实体机器人对人的认知情感的感知能力,以及社交网络上虚拟机器人的相关研究。国外的研究则更聚焦于使用者的感受和体验相关的因素研究。国内的教育机器人研究目前大多集中于STEAM (science, technology, engineering, art, and math)教育、编程教育和机器人教育,而国外的研究中教学的科目更为广泛,并不局限于机器人技术相关的教学。相较于国外远程呈现机器人的蓬勃发展,国内的研究总体较少。随着人们对社交机器人和教育机器人的需求不断增长,该领域或将迎来下一波重点关注。

3.3 个性化人机对话交互

在语音识别、声纹识别和语音合成领域,国内外研究方向趋同,对低资源及自适应语音识别、跨渠道声纹识别以及个性化语音合成等领域都有广泛关注。整体来讲,国外研究机构包括美国卡内基麦隆大学、新加坡南洋理工大学、英国剑桥大学等科研机构以及 Google、Facebook 等企业,以实际应用为导向,在原创性框架上做出了较多工作。国内学术界包括中国科学院自动化研究所、清华大学、西北工业大学等,企业包括科大讯飞、思必驰以及百度等,更侧重于语音识别、合成的技术落地,在性能上与国外无明显差距,并多次在国际相关比赛中取得佳绩,但在一些原创性的理论、方法上略有欠缺。在对话系统方面,国内外的研究大致同步,研究较集中在开放域对话,对口语理解中的情感因素、意图估计、知识模型以及个性化建模方面展开了广泛研究。

3.4 人机交互中的数据变换与可视化

表2给出了数据变换与可视化不同类技术之间的侧重点。其中,★表示关联性较弱,★★表示关联性适中,★★★表示关联性较强。尽管国内在人机交互、可视化方面近年有了很大发展,但总体与发达国家还有很大差距,特别是在人机交互的数据变换与可视化领域,国内研究尚处于萌芽阶段。2011年起,国外开始探索将推荐算法结合人机交互辅助数据变换,而国内尚未有相关研究成果发表。国内领先的BI软件具有灵活的交互式数据变换能力、丰富的可视化组件,但数据变换操作缺乏智能化。在数据变换可视化、数据表可视对比及代码可视化方面,国外研究侧重可视化数据内容本身,国内研究侧重数据变换语义的可视化。

表2 数据变换与可视化不同类别技术之间的侧重点对比

Table 2 Comparison of tasks among techniques in data transformation and visualization

可视化技术	交互式数据变换	数据变换可视化	数据表可视对比	代码可视化
展示数据表内容	★	★	★★	-
展示数据表差异	-	★★	★★★	-
展示数据变换语义	-	★★★	★★	★★
执行数据变换操作	★★★	-	-	★

注:“-”表示无法实现,★越多代表呈现度越高。

4 发展趋势与展望

4.1 人机交互中的感知计算

随着智能设备和传感器的广泛普及,利用感知计算提升用户交互体验将成为一种趋势。现有工作虽然已在众多真实世界的不同场景和任务中取得了较好成果,但对多设备的联合感知相对匮乏。同时,大部分感知研究都利用泛化的模型对人类行为进行感知,而忽视了个体之间的差异性,造成感知精度较低,无法在后期通过设计合理的交互对行为进行干预,在实际应用难以落地,另外,众多感知计算的研究都依赖于标记数据的丰富性和准确性,但如何针对不同的场景、任务和群体,进行基于迁移学习的感知计算也是研究难点。综上,未来感知计算的研究趋势是多模态、可迁移、个性化和规模化。

4.2 人与机器人交互及协同

随着工业化4.0的进程发展,人工智能技术对机器人产业的渗透带来了机器人进入人类社会的契机,对人与机器人的交互及协同产生了前所未有的需求和挑战。机器人产业作为硬科技的代表,具有前期研发投入大,研发周期长的特点,如果在研发前期没有能充分考虑人机交互接口设计,在后期用户体验时的调整则意味着巨大的改版成本。然而,相较于国外人与机器人相关研究百花齐放蓬勃发展的态势,目前国内的机器人研究并没有将人与机器人交互与协同作为一个重要的研究方向,特别是用户深度理解、交互数据集构建以及实际部署的长期用户体验研究方面,未来还需要系统性加深我国关于设计能够与人交互及协同的机器人的方法论构建。

4.3 个性化人机对话交互

语音识别、语音合成技术较为成熟,已在诸多产品中得到应用。目前学术界和产业界一般以应用为导引,结合具体人机交互具体场景,开展低资源、低延迟以及个性化方面的研究。尤其是语音合成方面,在音色、韵律方面更加重视个性化的研究。声纹识别或说话人认证旨在鉴定说话人身份,用以保护交互场景的安全,在智能家居、智能驾舱等产品中与语音识别联合应用。相关研究集中在跨渠道、低质和高噪声等较有挑战性的复杂场景,以进一步提高交互准确率。此外,结合预训练大模型(Fan等, 2021b),提取声学表征在语音识别、声纹识别是近两

年来较有潜力的发展方向。对话系统方面的研究多集中在开放域对话系统,结合预训练模型(Kim等, 2021a; Kim等, 2021b; Lai等, 2021),并试图对情感、意图和知识等方面提高建模精度,以实现人机对话更加智能化的目标。整体而言,在人机语音对话系统中,根据用户的具体特征针对性地涉及全部或部分系统,以适应用户的个性化需求,受到了越来越多的关注。

4.4 人机交互中的数据变换与可视化

人机交互中的数据变换与可视化通过人机界面直观呈现复杂大数据,使数据工作者交互式完成数据变换,提高工作效率。未来的研究方向可以分成两大类。首先,针对单个数据工作者,如何在交互中提高数据变换的智能化水平将会是研究热点。其中包括如何针对不同模态数据推荐相应的数据变换算法、如何根据用户的交互行为持续推荐数据变换与可视化方法以及如何支持海量数据的实时探索式分析等。其次,随着数据分析任务复杂度越来越高,如何支持多人协同变换数据、协同可视化是重要的研究方向。其中包括如何设计协同机制,消解不同数据工作者操作的冲突、如何解析并可可视化各类复杂数据变换代码,帮助他人理解数据变换语义以及如何评估种类繁多的数据变换可视化的效果,并自动推荐最适合的可视化方法、如何记录并展示不同人的分析行为,以提高协作效率等。

致谢 本文由中国图象图形学学会人机交互专业委员会组织撰写,该专委会链接为<http://www.csig.org.cn/detail/2490>。

参考文献(References)

- Abbou C C, Hoznek A, Salomon L, Olsson L E, Lobontiu A, Saint F, Cicco A, Antiphon P and Chopin D. 2017. Laparoscopic radical prostatectomy with a remote controlled robot. *The Journal of Urology*, 197(2S): S210-S212 [DOI: 10.1016/j.juro.2016.10.107]
- Abedjan Z, Morcos J, Ilyas I F, Ouzzani M, Papotti P and Stonebraker M. 2016. DataXFormer: a robust transformation discovery system// *Proceedings of the 32nd IEEE International Conference on Data Engineering (ICDE)*. Helsinki, Finland: IEEE: 1134-1145 [DOI: 10.1109/ICDE.2016.7498319]
- Adib F and Katabi D. 2013. See through walls with WiFi//*Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*. Hong Kong, China: ACM: 75-86 [DOI: 10.1145/2486001.2486039]
- Ajoudani A, Fang C, Tzagarakis N and Bicchi A. 2018. Reduced-complexity representation of the human arm active endpoint stiffness for supervisory control of remote manipulation. *The International Journal of Robotics Research*, 37(1): 155-167 [DOI: 10.1177/0278364917744035]
- An S M, Ling Z H and Dai L R. 2017. Emotional statistical parametric speech synthesis using LSTM-RNNS//*Proceedings of 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. Kuala Lumpur, Malaysia: IEEE: 1613-1616 [DOI: 10.1109/APSIPA.2017.8282282]
- Arief-Ang I B, Hamilton M and Salim F D. 2018. A scalable room occupancy prediction with transferable time series decomposition of CO₂ sensor data. *ACM Transactions on Sensor Networks (TOSN)*, 14(3/4): #21 [DOI: 10.1145/3217214]
- Arshad S, Feng C H, Liu Y H, Hu Y P, Yu R Y, Zhou S W and Li H. 2017. Wi-chase: a WiFi based human activity recognition system for sensorless environments//*Proceedings of the 18th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. Macau, China: IEEE: #7974315 [DOI: 10.1109/WoWMoM.2017.7974315]
- Atal B S. 1974. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *Journal of the Acoustical Society of America*, 55(6): 1304-1312 [DOI: 10.1121/1.1914702]
- Baevski A, Hsu W N, Conneau A and Auli M. 2021. Unsupervised speech recognition//*Proceedings of the 35th Conference on Neural Information Processing Systems*. Montreal, Canada: Curran Associates, Inc.: 27826-27839
- Bainbridge W A, Hart J W, Kim E S and Scassellati B. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1): 41-52 [DOI: 10.1007/s12369-010-0082-7]
- Balogh G and Beszédes Á. 2013. CodeMetropolis — A minecraft based collaboration tool for developers//*Proceedings of the 1st IEEE Working Conference on Software Visualization (VISSOFT)*. Eindhoven, Netherlands: IEEE: #6650528 [DOI: 10.1109/VISSOFT.2013.6650528]
- Bazzano F and Lamberti F. 2018. Human-robot interfaces for interactive receptionist systems and wayfinding applications. *Robotics*, 7(3): #56 [DOI: 10.3390/robotics7030056]
- Belpaeme T, Kennedy J, Ramachandran A, Scassellati B and Tanaka F. 2018. Social robots for education: a review. *Science Robotics*, 3(21): #eaat5954 [DOI: 10.1126/scirobotics.aat5954]
- Bhattacharjee T, Gordon E K, Scalise R, Cabrera M E, Caspi A, Cakmak M and Srinivasa S S. 2020. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding//*Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction*. Cambridge, UK: ACM: 181-190 [DOI: 10.1145/3319502.3374818]
- Bigelow A, Nobre C, Meyer M and Lex A. 2019. Origraph: interactive

- network wrangling//Proceedings of 2019 IEEE Conference on Visual Analytics Science and Technology (VAST). Vancouver, Canada: 81-92 [DOI: 10.1109/VAST47406.2019.8986909]
- Bogomolov A, Lepri B, Staiano J, Oliver N, Pianesi F and Pentland A. 2014. Once upon a crime: towards crime prediction from demographics and mobile data//Proceedings of the 16th International Conference on Multimodal Interaction. Istanbul, Turkey: ACM: 427-434 [DOI: 10.1145/2663204.2663254]
- Breazeal C L. 2000. *Sociable Machines: Expressive Social Exchange between Humans and Robots*. Cambridge, USA: Massachusetts Institute of Technology
- Casner S M, Hutchins E L and Norman D. 2016. The challenges of partially automated driving. *Communications of the ACM*, 59 (5) : 70-77 [DOI: 10.1145/2830565]
- Chen J. 2021. Study on Cyber Physical Modeling and Control Method of Intelligent Vehicle in Human Vehicle Co-piloting. Chongqing: Chongqing University (陈进. 2021. 智能汽车人机共驾信息物理建模及控制方法研究. 重庆: 重庆大学) [DOI: 10.27670/d.cnki.gcqdu.2021.003802]
- Chen L, Chang C, Chen Z, Tan B W, Gasic M and Yu K. 2018. Policy adaptation for deep reinforcement learning-based dialogue management//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Canada: IEEE: 6074-6078 [DOI: 10.1109/ICASSP. 2018. 8462272]
- Chen N X, Watanabe S, Villalba J, Żelasko P and Dehak N. 2020a. Non-autoregressive transformer for speech recognition. *IEEE Signal Processing Letters*, 28: 121-125 [DOI: 10.1109/LSP. 2020. 3044547]
- Chen R, Weng D, Huang Y W, Shu X H, Zhou J Y, Sun G D and Wu Y C. 2023. Rigel: transforming tabular data by declarative mapping. *IEEE Transactions on Visualization and Computer Graphics*, 29(1): 128-138 [DOI: 10.1109/TVCG.2022.3209385]
- Chen X L. 2022. Study on Teleoperation Control Method of Flexible Needle Puncture Assisted by Pneumatic Surgical Robot. Xuzhou: China University of Mining and Technology (陈肖利. 2022. 气动手术机器人辅助柔性针刺的遥操作控制方法研究. 徐州: 中国矿业大学) [DOI: 10.27623/d.cnki.gzkyu.2022.000485]
- Chen Z Y, Wang S and Qian Y M. 2020b. Adversarial domain adaptation for speaker verification using partially shared network//Proceedings of the 20th Annual Conference of the International Speech Communication Association. Shanghai, China: [s.n.]: 3017-3021 [DOI: 10.21437/Interspeech.2020-2226]
- Cheon J, Kang D and Woo G. 2015. VizMe: an annotation-based program visualization system generating a compact visualization//Proceedings of the International Conference on Data Engineering 2015 (DaEng-2015). Singapore, Singapore: Springer: 433-441 [DOI: 10.1007/978-981-13-1799-6_45]
- Chi E A, Salazar J and Kirchoff K. 2020. Align-Refine: non-autoregressive speech recognition via iterative realignment [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/2010.14233.pdf>
- Chu M D, Zong K Y, Shu X, Gong J T, Lu Z C, Guo K M, Dai X Y and Zhou G Y. 2023. Work with AI and work for AI: autonomous vehicle safety drivers' lived experiences//Proceedings of 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany. ACM: 1-16 [DOI: 10.1145/3544548.35815]
- Chung Y A, Wang Y X, Hsu W N, Zhang Y and Skerry-Ryan R J. 2019. Semi-supervised training for improving data efficiency in end-to-end speech synthesis//Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, UK: IEEE: 6940-6944 [DOI: 10.1109/ICASSP.2019.8683862]
- Claessen J H T and van Wijk J J. 2011. Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12): 2310-2316 [DOI: 10.1109/TVCG. 2011.201]
- Coghlan S, Waycott J, Lazar A and Neves B B. 2021. Dignity, autonomy, and style of company: dimensions older adults consider for robot companions. *Proceedings of the ACM on Human-Computer Interaction*, 5: #104 [DOI: 10.1145/3449178]
- Coradeschi S, Loutfi A, Kristoffersson A, Cortellessa G and Eklundh K S. 2011. Social robotic telepresence//Proceedings of the 6th International Conference on Human-Robot Interaction. Lausanne, Switzerland: Association for Computing Machinery: #1957660 [DOI: 10.1145/1957656.1957660]
- Cross E S, Hortensius R and Wykowska A. 2019. From social brains to social robots: applying neurocognitive insights to human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374 (1771) : #20180024 [DOI: 10.1098/rstb. 2018.0024]
- Dai Y H, Chen T Y, Zhi T and Li Q Y. 2022. Design and implementation of service robot and existing system facilities fusion application in high speed railway station. *Railway Transport and Economy*, 44(9): 77-82 (戴彦华, 陈天煜, 支涛, 李全印. 2022. 服务机器人与铁路客运站既有系统设施融合应用的设计与实现. 铁道运输与经济, 44(9): 77-82) [DOI: 10.16668/j.cnki.issn.1003-1421. 2022.09.11]
- Daruwalla Z J, Collins D R and Moore D P. 2010. "Orthobot, to your station!" The application of the remote presence robotic system in orthopaedic surgery in Ireland: a pilot study on patient and nursing staff satisfaction. *Journal of Robotic Surgery*, 4 (3) : 177-182 [DOI: 10.1007/s11701-010-0207-x]
- Davis S and Mermelstein P. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4): 357-366 [DOI: 10.1109/tassp.1980.1163420]
- Dehak N, Kenny P J, Dehak R, Dumouchel P and Ouellet P. 2011. Front-end factor analysis for speaker verification. *IEEE Transac-*

- tions on Audio, Speech, and Language Processing, 19 (4) : 788-798 [DOI: 10.1109/TASL.2010.2064307]
- Delcroix M, Watanabe S, Ogawa A, Karita S and Nakatani T. 2018. Auxiliary feature based adaptation of end-to-end ASR systems//Proceedings of 2018 Annual Conference of the International Speech Communication Association. Hyderabad, India: [s. n.] : 2444-2448 [DOI: 10.21437/Interspeech.2018-1438]
- Demetrescu C, Finocchi I and Stasko J T. 2002. Specifying algorithm visualizations: interesting events or state mapping?//Software Visualization. Castle, Germany: Springer: 16-30 [DOI: 10.1007/3-540-45875-1_2]
- Dey A K, Salber D, Abowd G D and Futakawa M. 1999. The conference assistant: combining context-awareness with wearable computing//Proceedings of Digest of Papers. The 3rd International Symposium on Wearable Computers. San Francisco, USA: IEEE: 21-28 [DOI: 10.1109/ISWC.1999.806639]
- Di Lascio E, Gashi S and Santini S. 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2 (3) : #103 [DOI: 10.1145/3264913]
- Drosos I, Barik T, Guo P J, DeLine R and Gulwani S. 2020. Wrex: a unified programming-by-example interaction for synthesizing readable code for data scientists//Proceedings of 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, USA: ACM: 1-12 [DOI: 10.1145/3313831.3376442]
- Fan Z Y, Li J, Zhou S Y and Xu B. 2019. Speaker-aware speech-transformer//Proceedings of 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). Singapore, Singapore: IEEE: 222-229 [DOI: 10.1109/ASRU46091.2019.9003844]
- Fan Z Y, Li M, Zhou S Y and Xu B. 2021b. Exploring wav2vec 2.0 on speaker verification and language identification//Proceedings of 2012 Annual Conference of the International Speech Communication Association. Brno, Czechia: [s.n.]: 1509-1513
- Fan Z Y, Liang Z L, Dong L H, Liu Y, Zhou S Y, Cai M, Zhang J, Ma Z J and Xu B. 2022. Token-level speaker change detection using speaker difference and speech content via continuous integrate-and-fire//Proceedings of 2022 Annual Conference of the International Speech Communication Association. Incheon, Korea (South) : [s. n.]: 3749-3753
- Fan Z Y, Zhou S Y and Xu B. 2021a. Two-stage pre-training for sequence to sequence speech recognition//Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen, China: IEEE: #9534170 [DOI: 10.1109/IJCNN52387.2021.9534170]
- Fang B, Guo D, Sun F C, Liu H P and Wu Y P. 2015. A robotic hand-arm teleoperation system using human arm/hand with a novel data glove//Proceedings of 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO). Zhuhai, China: IEEE: 2483-2488 [DOI: 10.1109/ROBIO.2015.7419712]
- Figueredo L F C, De Castro Aguiar R, Chen L P, Richards T C, Chakrabarty S and Dogar M. 2021. Planning to minimize the human muscular effort during forceful human-robot collaboration. ACM Transactions on Human-Robot Interaction, 11(1) : #10 [DOI: 10.1145/3481587]
- Finn C, Abbeel P and Levine S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 1126-1135
- Foster M E. 2019. Natural language generation for social robotics: Opportunities and challenges. Philosophical Transactions of the Royal Society B: Biological Sciences, 374 (1771) : #20180027 [DOI: 10.1098/rstb.2018.0027]
- Franzluebbers A and Johnson K. 2019. Remote robotic arm teleoperation through virtual reality//Proceedings of 2019 Symposium on Spatial User Interaction. New Orleans, USA: Association for Computing Machinery: #27 [DOI: 10.1145/3357251.3359444]
- Fridman L. 2018. Human-centered autonomous vehicle systems: principles of effective shared autonomy [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/1810.01835.pdf>
- Fua Y H, Ward M O and Rundensteiner E A. 1999. Hierarchical parallel coordinates for exploration of large datasets//Proceedings of the Visualization' 99. San Francisco, USA: IEEE: 43-50 [DOI: 10.1109/VISUAL.1999.809866]
- Furmanova K, Gratzl S, Stütz H, Zichner T, Jaresova M, Lex A and Streit M. 2020. Taggle: combining overview and details in tabular data visualizations. Information Visualization, 19 (2) : 114-136 [DOI: 10.1177/1473871619878085]
- Gales M and Young S. 2008. The application of hidden Markov models in speech recognition. Foundations and Trends[®] in Signal Processing, 1(3) : 195-304 [DOI: 10.1561/20000000004]
- Galin R and Meshcheryakov R. 2021. Collaborative robots: development of robotic perception system, safety issues, and integration of AI to imitate human behavior//Proceedings of 15th International Conference on Electromechanics and Robotics "Zavalishin's Readings". Ufa, Russia: Springer: 175-185 [DOI: 10.1007/978-981-15-5580-0_14]
- Gangadharaiah R and Narayanaswamy B. 2019. Joint multiple intent detection and slot labeling for goal-oriented dialog//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minnesota, USA: Association for Computational Linguistics: 564-569 [DOI: 10.18653/v1/N19-1055]
- Gao B J, Xu J J, Du J and Huang R H. 2020. Functional analysis framework and case study of educational robot products. Modern Educational Technology, 30(1) : 18-24 (高博俊, 徐晶晶, 杜静, 黄荣怀. 2020. 教育机器人产品的功能分析框架及其案例研究. 现代

- 教育技术, 30(1): 18-24 [DOI: 10.3969/j.issn.1009-8097.2020.01.003]
- Gao N, Rahaman M S, Shao W, Ji K X and Salim F D. 2022a. Individual and group-wise classroom seating experience: effects on student engagement in different courses. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(3): #115 [DOI: 10.1145/3550335]
- Gao N, Shao W, Rahaman M S and Salim F D. 2020b. *n-Gage*: predicting in-class emotional, behavioural and cognitive engagement in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3): #79 [DOI: 10.1145/3411813]
- Gao N, Shao W, Rahaman M S, Zhai J, David K and Salim F D. 2021. Transfer learning for thermal comfort prediction in multiple cities. *Building and Environment*, 195: #107725 [DOI: 10.1016/j.buildenv.2021.107725]
- Gao N, Shao W and Salim F D. 2019. Predicting personality traits from physical activity intensity. *Computer*, 52(7): 47-56 [DOI: 10.1109/MC.2019.2913751]
- Gao S L, Takanobu R, Bosselut A and Huang M L. 2022b. End-to-end task-oriented dialog modeling with semi-structured knowledge management. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30: 2173-2187 [DOI: 10.1109/TASLP.2022.3153255]
- Gao J S, Gong J T, Zhou G Y, Guo H L and Qi T. 2022c. Learning with yourself: a tangible twin robot system to promote STEM education. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4981-4988 [DOI: 10.1109/IROS47612.2022.9981423]
- Gauvain J L and Lee C H. 1992. Bayesian learning for hidden Markov model with Gaussian mixture state observation densities. *Speech Communication*, 11(2/3): 205-213 [DOI: 10.1016/0167-6393(92)90015-Y]
- Gauvain J L and Lee C H. 1994. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2): 291-298 [DOI: 10.1109/89.279278]
- Gorham J. 1988. The relationship between verbal teacher immediacy behaviors and student learning. *Communication Education*, 37(1): 40-53 [DOI: 10.1080/03634528809378702]
- Gouaillier D, Hugel V, Blazevic P, Kilner C, Monceaux J, Lafourcade P, Marnier B, Serre J and Maisonnier B. 2009. Mechatronic design of NAO humanoid//*Proceedings of 2009 IEEE International Conference on Robotics and Automation*. Kobe, Japan: IEEE: 769-774 [DOI: 10.1109/ROBOT.2009.5152516]
- Gratz S, Lex A, Gehlenborg N, Pfister H and Streit M. 2013. LineUp: visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12): 2277-2286 [DOI: 10.1109/TVCG.2013.173]
- Graves A, Mohamed A R and Hinton G. 2013. Speech recognition with deep recurrent neural networks//*Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, Canada: IEEE: 6645-6649 [DOI: 10.1109/ICASSP.2013.6638947]
- Guerreiro J, Sato D, Asakawa S, Dong H X, Kitani K M and Asakawa C. 2019. CaBot: designing and evaluating an autonomous navigation robot for blind people//*Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. Pittsburgh, USA: Association for Computing Machinery: 68-82 [DOI: 10.1145/3308561.3353771]
- Gunes H, Celiktutan O and Sariyanidi E. 2019. Live human-robot interactive public demonstrations with automatic emotion and personality prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771): #20180026 [DOI: 10.1098/rstb.2018.0026]
- Guo B, Yu Z W, Chen L M, Zhou X S and Ma X J. 2016. MobiGroup: enabling lifecycle support to social activity organization and suggestion with mobile crowd sensing. *IEEE Transactions on Human-Machine Systems*, 46(3): 390-402 [DOI: 10.1109/THMS.2015.2503290]
- Guo P C, Boyer F, Chang X K, Hayashi T, Higuchi Y, Inaguma H, Kamo N, Li C D, Garcia-Romero D, Shi J T, Shi J, Watanabe S, Wei K, Zhang W Y and Zhang Y K. 2021. Recent developments on Espnet toolkit boosted by conformer//*Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, Canada: IEEE: 5874-5878
- Guo P J. 2013. Online python tutor: embeddable web-based program visualization for CS education//*Proceedings of the 44th ACM Technical Symposium on Computer Science Education*. Denver, USA: ACM: 579-584 [DOI: 10.1145/2445196.2445368]
- Guo P J, Kandel S, Hellerstein J M and Heer J. 2011. Proactive wrangling: mixed-initiative end-user programming of data transformation scripts//*Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. Santa Barbara, USA: ACM: 65-74 [DOI: 10.1145/2047196.2047205]
- Han J Y. 2022. Research on Human-Machine Haptic Interactive Shared Steering Control Based on Driver Behavior Understanding for Intelligent Vehicle. Changchun: Jilin University (韩嘉懿. 2022. 基于驾驶人行为理解的智能汽车人机触觉交互协同转向控制研究. 长春: 吉林大学) [DOI: 10.27162/d.enki.gjlin.2022.001199]
- Handa A, Van Wyk K, Yang W, Liang J, Chao Y W, Wan Q, Birchfield S, Ratliff N and Fox D. 2020. DexPilot: vision-based teleoperation of dexterous robotic hand-arm system//*Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA)*. Paris, France: IEEE: 9164-9170 [DOI: 10.1109/ICRA40945.2020.9197124]
- Hansen S, Narayanan N H and Hegarty M. 2002. Designing educationally effective algorithm visualizations. *Journal of Visual Languages and Computing*, 13(3): 291-317 [DOI: 10.1006/jvlc.2002.0236]

- Hashimoto S, Ishida A, Inami M and Igarashi T. 2011. TouchMe: an augmented reality based remote robot manipulation//Proceedings of the 21st International Conference on Artificial Reality and Telexistence. Osaka, Japan: The Virtual Reality Society of Japan: 61-66
- Heigold G, Moreno I, Bengio S and Shazeer N. 2015. End-to-end text-dependent speaker verification//Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Shanghai, China: IEEE: 5115-5119 [DOI: 10.1109/ICASSP.2016.7472652]
- Hentout A, Aouache M, Maoudj A and Akli I. 2019. Human-robot interaction in industrial collaborative robotics: a literature review of the decade 2008-2017. *Advanced Robotics*, 33 (15/16): 764-799 [DOI: 10.1080/01691864.2019.1636714]
- Higuchi Y, Inaguma H, Watanabe S, Ogawa T and Kobayashi T. 2021. Improved mask-CTC for non-autoregressive end-to-end ASR//Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, Canada: IEEE: 8363-8367 [DOI: 10.1109/ICASSP39728.2021.9414198]
- Hood D, Lemaignan S and Dillenbourg P. 2015. When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting//Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction. Portland, USA: Association for Computing Machinery: 83-90 [DOI: 10.1145/2696454.2696479]
- Hori A, Kawakami M and Ichii M. 2019. CodeHouse: VR code visualization tool//Proceedings of 2019 Working Conference on Software Visualization (VISSOFT). Cleveland, USA: IEEE: 83-87 [DOI: 10.1109/VISSOFT.2019.00018]
- Hu H R, Song Y, Dai L R, McLoughlin I and Liu L. 2022. Class-aware distribution alignment based unsupervised domain adaptation for speaker verification//Interspeech 2022. Incheon, Korea (South): [s.n.]: 3689-3693 [DOI: 10.21437/Interspeech.2022-591]
- Hu S. 2021. The use of program visualization in the teaching of programming design courses. *Computer Knowledge and Techniques*, 17(7): 104-105 (胡珊. 2021. 程序可视化方法在程序设计课程教学中的应用. *电脑知识与技术*, 17(7): 104-105) [DOI: 10.14004/j.cnki.ckt.2021.0747]
- Hu Y. 2020. On Collision Detection Method of Human-Machine Cooperative Robot. Qingdao: Shandong University of Science and Technology (胡钰. 2020. 人机协作机器人的碰撞检测方法研究. 青岛: 山东科技大学) [DOI: 10.27275/d.cnki.gsdku.2020.001003]
- Huang J H, Floyd M F, Tateosian L G and Hipp J A. 2022. Exploring public values through Twitter data associated with urban parks pre and post- COVID-19. *Landscape and Urban Planning*, 227: #104517 [DOI: 10.1016/j.landurbplan.2022.104517]
- Huang W Y, Hu W C, Yeung Y T and Chen X. 2020. Conv-transformer transducer: low latency, low frame rate, streamable end-to-end speech recognition [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/2008.05750.pdf>
- Huang Z, Zhuang X D, Liu D B, Xiao X Q, Zhang Y C and Siniscalchi S M. 2019. Exploring retraining-free speech recognition for intra-sentential code-switching//Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, UK: IEEE: 6066-6070 [DOI: 10.1109/ICASSP.2019.8682478]
- Huynh S, Kim S, Ko J G, Balan R K and Lee Y. 2018. EngageMon: multi-modal engagement sensing for mobile games. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1): #13 [DOI: 10.1145/3191745]
- Inaguma H, Mimura M and Kawahara T. 2020. Enhancing monotonic multihead attention for streaming ASR [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/2005.09394.pdf>
- Inala J P and Singh R. 2017. WebRelate: integrating web data with spreadsheets using examples. *Proceedings of the ACM on Programming Languages*, 2: #2 [DOI: 10.1145/3158090]
- Itakura F. 1975. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23 (1): 67-72 [DOI: 10.1109/tassp.1975.1162641]
- Ivanov S H, Webster C and Berezina K. 2017. Adoption of robots and service automation by tourism and hospitality companies. *Revista Turismo and Desenvolvimento*, (27/28): 1501-1517
- Jain S and Argall B. 2020. Probabilistic human intent recognition for shared autonomy in assistive robotics. *ACM Transactions on Human-Robot Interaction*, 9(1): #2 [DOI: 10.1145/3359614]
- Jalal A, Kim Y H, Kim Y J, Kamal S and Kim D. 2017. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognition*, 61: 295-308 [DOI: 10.1016/j.patcog.2016.08.003]
- Jbara A, Agbaria M, Adoni A, Jabareen M and Yasin A. 2019. ICSD: interactive visual support for understanding code control structure//Proceedings of the 26th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). Hangzhou, China: IEEE: 644-648 [DOI: 10.1109/SANER.2019.8667981]
- Jiao Z Q, Meng Q L, Shao H C and Yu H L. 2022. Design and research of the upper limb rehabilitation training and assisted living robot. *Chinese Journal of Rehabilitation Medicine*, 37 (9): 1219-1222 (焦宗琪, 孟巧玲, 邵海存, 喻洪流. 2022. 上肢康复训练与生活辅助机器人的设计与研究. *中国康复医学杂志*, 37(9): 1219-1222) [DOI: 10.3969/j.issn.1001-1242.2022.09.011]
- Jin Z J, Anderson M R, Cafarella M and Jagadish H V. 2017. Foofah: transforming data by example//Proceedings of 2017 ACM International Conference on Management of Data. Chicago, USA: ACM: 683-698 [DOI: 10.1145/3035918.3064034]
- Jobanputra C, Bavishi J and Doshi N. 2019. Human activity recognition: a survey. *Procedia Computer Science*, 155: 698-703 [DOI: 10.1016/j.procs.2019.08.100]
- Johnsson S L and Krawitz R L. 1992. Cooley-Tukey FFT on the connec-

- tion machine. *Parallel Computing*, 18(11): 1201-1221 [DOI: 10.1016/0167-8191(92)90066-G]
- Kachouie R, Sedighadeli S M, Khosla R and Chu M T. 2014. Socially assistive robots in elderly care: a mixed-method systematic literature review. *International Journal of Human-Computer Interaction*, 30(5): 369-393 [DOI: 10.1080/10447318.2013.873278]
- Kahn J, Lee A and Hannun A. 2020. Self-training for end-to-end speech recognition//*Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE: 7084-7088 [DOI: 10.1109/ICASSP40776.2020.9054295]
- Kandel S, Paepcke A, Hellerstein J and Heer J. 2011. Wrangler: interactive visual specification of data transformation scripts//*Proceedings of 2011 SIGCHI Conference on Human Factors in Computing Systems*. Vancouver, Canada: ACM: 3363-3372 [DOI: 10.1145/1978942.1979444]
- Karita S, Watanabe S, Iwata T, Ogawa A and Delcroix M. 2018. Semi-supervised end-to-end speech recognition//*Proceedings of 2018 Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Hyderabad, India: ISCA: 2-6 [DOI: 10.21437/Interspeech.2018-1746]
- Kennedy J, Baxter P and Belpaeme T. 2015. Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, 7(2): 293-308 [DOI: 10.1007/s12369-014-0277-4]
- Kenny P, Boulianne G, Ouellet P and Dumouchel P. 2007. Joint factor analysis versus eigenchannels in speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4): 1435-1447 [DOI: 10.1109/TASL.2006.881693]
- Khadri H O. 2021. University academics' perceptions regarding the future use of telepresence robots to enhance virtual transnational education: an exploratory investigation in a developing country. *Smart Learning Environments*, 8(1): #28 [DOI: 10.1186/s40561-021-00173-8]
- Khaloo P, Maghomi M, Taranta E, Bettner D and Laviola J. 2017. Code park: a new 3D code visualization tool//*Proceedings of 2017 IEEE Working Conference on Software Visualization (VISSOFT)*. Shanghai, China: IEEE: 43-53 [DOI: 10.1109/VISSOFT.2017.10]
- Khan M, Xu L, Nandi A and Hellerstein J M. 2017. Data tweening: incremental visualization of data transforms. *Proceedings of the VLDB Endowment*, 10(6): 661-672 [DOI: 10.14778/3055330.3055333]
- Kidd C D and Breazeal C. 2004. Effect of a robot on user perceptions//*Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Sendai, Japan: IEEE: 3559-3564 [DOI: 10.1109/IROS.2004.1389967]
- Kim M, Kim G, Lee S W and Ha J W. 2021a. St-Bert: cross-modal language model pre-training for end-to-end spoken language understanding//*Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, Canada: IEEE: 7478-7482 [DOI: 10.1109/ICASSP39728.2021.9414558]
- Kim S, Kim G, Shin S and Lee S. 2021b. Two-stage textual knowledge distillation for end-to-end spoken language understanding [EB/OL]. [2021-06-10]. <https://arxiv.org/pdf/2010.13105.pdf>
- Köse H, Uluer P, Akalın N, Yorgancı R, Özkul A and Ince G. 2015. The effect of embodiment in sign language tutoring with assistive humanoid robots. *International Journal of Social Robotics*, 7(4): 537-548 [DOI: 10.1007/s12369-015-0311-1]
- Kosower D A, Lopez-Villarejo J J and Roubtsov S. 2014. Flowgen: flowchart-based documentation framework for C++//*Proceedings of the 14th IEEE International Working Conference on Source Code Analysis and Manipulation*. Victoria, Canada: IEEE: 59-64 [DOI: 10.1109/SCAM.2014.35]
- Kristoffersson A, Coradeschi S, and Loutfi A. 2013. A review of mobile robotic telepresence. *Adv. in Hum.-Comp. Int.*: #3 [DOI: 10.1155/2013/902316]
- Kuo C M, Chen L C, Tseng C Y. 2017. Investigating an innovative service with hospitality robots. *International Journal of Contemporary Hospitality Management*, 29(5): 1305-1321
- Kumar N S, Revanth Babu P N, Sai Eashwar K S, Srinath M P and Bothra S. 2021. Code-Viz: data structure specific visualization and animation tool for user-provided code//*Proceedings of 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. Pune, India: IEEE: #9645747 [DOI: 10.1109/SMARTGENCON51891.2021.9645747]
- Lai C I, Chuang Y S, Lee H Y, Li S W and Glass J. 2021. Semi-supervised spoken language understanding via self-supervised speech and language model pretraining//*Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, Canada: IEEE: 7468-7472 [DOI: 10.1109/ICASSP39728.2021.9414922]
- Lan O Y, Zhu S and Yu K. 2018. Semi-supervised training using adversarial multi-task learning for spoken language understanding//*Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, Canada: IEEE: 6049-6053 [DOI: 10.1109/ICASSP.2018.8462669]
- Lei Y, Yang S, Cong J, Xie L and Su D. 2022. Glow-WaveGAN 2: high-quality zero-shot text-to-speech synthesis and any-to-any voice conversion [EB/OL]. [2022-07-05]. <https://arxiv.org/pdf/2207.01832.pdf>
- Leite I, Pereira A, Castellano G, Mascarenhas S, Martinho C and Paiva A. 2011. Social robots in learning environments: a case study of an empathic chess companion//*Proceedings of 2011 International Workshop on Personalization Approaches in Learning Environments*. Girona, Spain: CEUR-WS: 8-12
- Lex A, Schulz H J, Streit M, Partl C and Schmalstieg D. 2011. Vis-

- Bricks: multiform visualization of large, inhomogeneous data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12): 2291-2300 [DOI: 10.1109/TVCG.2011.250]
- Lex A, Streit M, Schulz H J, Partl C, Schmalstieg D, Park P J and Gehlenborg N. 2012. StratomeX: visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization. *Computer Graphics Forum*, 31: 1175-1184 [DOI: 10.1111/j.1467-8659.2012.03110.x]
- Leyzberg D, Spaulding S, Toneva M and Scassellati B. 2012. The physical presence of a robot tutor increases cognitive learning gains//Proceedings of the 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan: the Cognitive Science Society: 1882-1887
- Li G Z, Li R F, Wang Z C, Liu C H, Lu M and Wang G R. 2023. HiTailor: interactive transformation and visualization for hierarchical tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 29(1): 139-148 [DOI: 10.1109/TVCG.2022.3209354]
- Li J. 2015. The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77: 23-37 [DOI: 10.1016/j.ijhcs.2015.01.001]
- Li J, Fang X, Chu F, Gao T, Song Y and Dai R L. 2022a. Acoustic feature shuffling network for text-independent speaker verification//Proceedings of Interspeech 2022. Incheon, Korea (South): [s.n.]: 4790-4794 [DOI: 10.21437/Interspeech.2022-10278]
- Li J B, Meng Y, Wu X X, Wu Z Y, Jia J, Meng H L, Tian Q, Wang Y P and Wang Y X. 2022b. Inferring speaking styles from multimodal conversational context by multi-scale relational graph convolutional networks//Proceedings of the 30th ACM International Conference on Multimedia. Lisbon, Portugal: ACM: 5811-5820 [DOI: 10.1145/3503161.3547831]
- Li N. 2018. Research and Implementation on Path Planning of Industrial Robot towards Safety Assurance of Human-Robot Collaboration. Wuhan: Wuhan University of Technology (李娜. 2018. 面向人机协作安全保障的工业机器人路径规划研究与实现. 武汉: 武汉理工大学)
- Li P L. 2018. An Interactive Method for Telepresence System Based on Stereo Vision and Gesture Recognition. Beijing: Beijing Institute of Technology (李佩霖. 2018. 一种基于立体视觉与手势识别的远程呈现交互方法. 北京: 北京理工大学) [DOI: 10.26948/d.cnki.gbjlu.2018.000433]
- Li Z X. 2022. Research on Intelligent Assistance Method of Nursing Robot in the Stand-to-sit Interaction. Shenyang: Shenyang University of Technology (李泽新. 2022. 站坐交互过程中护理机器人智能辅助方法研究. 沈阳: 沈阳工业大学) [DOI: 10.27322/d.cnki.gsgyu.2022.000214]
- Liang C, Yu C, Qin Y, Wang Y T and Shi Y C. 2021. DualRing: enabling subtle and expressive hand interaction with dual IMU rings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3): #115 [DOI: 10.1145/3478114]
- Liao H. 2013. Speaker adaptation of context dependent deep neural networks//Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada: IEEE: 7947-7951 [DOI: 10.1109/ICASSP.2013.6639212]
- Lin Y P, Wang C H, Jung T P, Wu T L, Jeng S K, Duann J R and Chen J H. 2010. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7): 1798-1806 [DOI: 10.1109/TBME.2010.2048568]
- Liu F, Mao Q R, Wang L J, Ruwa N, Gou J P and Zhan Y Z. 2019. An emotion-based responding model for natural language conversation. *World Wide Web*, 22(2): 843-861 [DOI: 10.1007/s11280-018-0601-2]
- Liu S. 2022. Design of Intelligent Service Robot based on Situation Model. Jinan: Shandong Jianzhu University (刘双. 2022. 基于情境模型的智能服务机器人设计研究. 济南: 山东建筑大学) [DOI: 10.27273/d.cnki.gsjc.2022.000213]
- Liu S S, Maljovec D, Wang B, Bremer P T and Pascucci V. 2017. Visualizing high-dimensional data: advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3): 1249-1268 [DOI: 10.1109/TVCG.2016.2640960]
- Liu T T, Liu Z, Chai Y J, Wang J and Wang Y Y. 2021. Agent affective computing in human-computer interaction. *Journal of Image and Graphics*, 26(12): 2767-2777 (刘婷婷, 刘箴, 柴艳杰, 王瑾, 王媛怡. 2021. 人机交互中的智能体情感计算研究. 中国图象图形学报, 26(12): 2767-2777) [DOI: 10.11834/jig.200498]
- Liu Z C, Wu N Q, Zhang Y J and Ling Z H. 2022. Integrating discrete word-level style variations into non-autoregressive acoustic models for speech synthesis//Proceedings of Interspeech 2022. Incheon, Korea (South): [s.n.]: 5508-5512 [DOI: 10.21437/Interspeech.2022-984]
- Long M L and Kang H Y. 2022. Malicious code detection for industrial internet based on code visualization. *Computer-Integrated Manufacturing Systems*: 1-14 (龙墨澜, 康海燕. 2022. 基于代码可视化的工业互联网恶意代码检测方法. 计算机集成制造系统, 1-14)
- Lorenzo-Trueba J, Henter G E, Takaki S, Yamagishi J, Morino Y and Ochiai Y. 2018. Investigating different representations for modeling and controlling multiple emotions in DNN-based speech synthesis. *Speech Communication*, 99: 135-143 [DOI: 10.1016/j.specom.2018.03.002]
- Lu L, Cai R Y and Gursoy D. 2019. Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management*, 80: 36-51 [DOI: 10.1016/j.ijhm.2019.01.005]
- Lu W, Wang J D, Chen Y Q, Pan S J, Hu C Y and Qin X. 2022. Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2): #65 [DOI: 10.1145/3534589]
- Luck J E. 1969. Automatic speaker verification using cepstral measure-

- ments. *Journal of the Acoustical Society of America*, 46(4): #1026 [DOI: 10.1121/1.1911795]
- Luo X N, Yuan Y, Zhang K Y, Xia J Z, Zhou Z G, Chang L and Gu T L. 2019. Enhancing statistical charts: toward better data visualization and analysis. *Journal of Visualization*, 22(4): 819-832 [DOI: 10.1007/s12650-019-00569-2]
- Müller M. 2007. *Dynamic time warping//Information Retrieval for Music and Motion*. Berlin, Heidelberg: Springer: 69-84 [DOI: 10.1007/978-3-540-74048-3_4]
- Ma H T. 2022. *Safety Analysis and Evaluation of Human-Machine Interaction of Automated Lane Keeping System*. Jilin: Jilin University (马海涛. 2022. 自动车道保持系统人机交互安全分析与评价. 吉林: 吉林大学) [DOI: 10.27162/d.cnki.gjlin.2022.006085]
- Ma N and Yuan X R. 2020. Tabular data visualization interactive construction for analysis tasks. *Journal of Computer-Aided Design and Computer Graphics*, 32(10): 1628-1636 (马楠, 袁晓如. 2020. 面向分析任务的表格数据可视化交互构建. *计算机辅助设计与图形学学报*, 32(10): 1628-1636) [DOI: 10.3724/SP.J.1089.2020.18467]
- Madotto A, Wu C S and Fung P. 2018. Mem2Seq: effectively incorporating knowledge bases into end-to-end task-oriented dialog systems. *arXiv preprint [EB/OL]*. [2023-01-13]. <https://arxiv.org/pdf/1804.08217.pdf>
- Marchetti E, Grimme S, Hornecker E, Kollakidou A and Graf P. 2022. Pet-robot or appliance? Care home residents with dementia respond to a zoomorphic floor washing robot//*Proceedings of 2022 CHI Conference on Human Factors in Computing Systems*. New Orleans, USA: Association for Computing Machinery: #521 [DOI: 10.1145/3491102.3517463]
- Min S, Lee B and Yoon S. 2017. Deep learning in bioinformatics. *Briefings in Bioinformatics*, 18(5): 851-869 [DOI: 10.1093/bib/bbw068]
- Mohamed A R, Dahl G and Hinton G. 2010. Deep belief networks for phone recognition//*Proceedings of NIPS Workshop on Deep Learning for Speech Recognition and Related Applications*. [s.l.]: [s.n.]: #39
- Moseler O, Kreber L and Diehl S. 2022. The ThreadRadar visualization for debugging concurrent Java programs. *Journal of Visualization*, 25(6): 1267-1289 [DOI: 10.1007/s12650-022-00843-w]
- Mu W B, Xu B Y, Guo W T, Wahafu T, Zou C, Ji B C and Cao L. 2022. Early clinical outcomes of MAKO robot-assisted total knee arthroplasty for knee osteoarthritis with severe varus deformity. *Chinese Journal Bone and Joint Surgery*, 15(8): 612-618 (穆文博, 胥伯勇, 郭文涛, 吐尔洪江·瓦哈甫, 邹晨, 纪保超, 曹力. 2022. MAKO机器人辅助下全膝关节置换术治疗伴有严重内翻畸形膝关节炎的早期疗效研究. *中华骨与关节外科杂志*, 15(8): 612-618) [DOI: 10.3969/j.issn.2095-9958.2022.08.09]
- Murphy J, Gretzel U and Pesonen J. 2019. Marketing robot services in hospitality and tourism: the role of anthropomorphism. *Journal of Travel and Tourism Marketing*, 36(7): 784-795 [DOI: 10.1080/10548408.2019.1571983]
- Niederer C, Stitz H, Hourieh R, Grassinger F, Aigner W and Streit M. 2018. TACO: visualizing changes in tables over time. *IEEE Transactions on Visualization and Computer Graphics*, 24(1): 677-686 [DOI: 10.1109/TVCG.2017.2745298]
- Niu K, Zhang F S, Chang Z X and Zhang D Q. 2018. A Fresnel diffraction model based human respiration detection system using COTS Wi-Fi devices//*Proceedings of 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. Singapore, Singapore: ACM: 416-419 [DOI: 10.1145/3267305.3267561]
- Obuchi M, Huckins J F, Wang W C, daSilva A, Rogers C, Murphy E, Hedlund E, Holtzheimer P, Mirjafari S and Campbell A. 2020. Predicting brain functional connectivity using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1): #23 [DOI: 10.1145/3381001]
- Ochiai T, Watanabe S, Katagiri S, Hori T and Hershey J. 2018. Speaker adaptation for multichannel end-to-end speech recognition//*Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, Canada: IEEE: 6707-6711 [DOI: 10.1109/ICASSP.2018.8462161]
- Ogawa K, Nishio S, Koda K, Taura K, Minato T, Ishii C T and Ishiguro H. 2011. Telenoid: tele-presence android for communication//*Proceedings of ACM SIGGRAPH 2011 Emerging Technologies*. Vancouver, Canada: Association for Computing Machinery: #15 [DOI: 10.1145/2048259.2048274]
- Osawa H, Ema A, Hattori H, Akiya N, Kanzaki N, Kubo A, Koyama T and Ichise R. 2017. Analysis of robot hotel: reconstruction of works with robots//*Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Lisbon, Portugal: IEEE: 219-223 [DOI: 10.1109/ROMAN.2017.8172305]
- Pajer S, Streit M, Torsney-Weir T, Spechtenhauser F, Möller T and Piringier H. 2017. WeightLifter: visual weight space exploration for multi-criteria decision making. *IEEE Transactions on Visualization and Computer Graphics*, 23(1): 611-620 [DOI: 10.1109/TVCG.2016.2598589]
- Pakarinen T, Pietilä J and Nieminen H. 2019. Prediction of self-perceived stress and arousal based on electrodermal activity//*Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany: IEEE: 2191-2195 [DOI: 10.1109/EMBC.2019.8857621]
- Pang G F. 2022. *Research on Interaction Design of Home Elderly Service Robot based on User Experience*. Yangzhou: Yangzhou University (庞广风. 2022. 基于用户体验的家庭助老服务机器人交互设计研究. 扬州: 扬州大学) [DOI: 10.27441/d.cnki.gyzdu.2022.002178]

- Paul P and George T. 2015. An effective approach for human activity recognition on smartphone//Proceedings of 2015 IEEE International Conference on Engineering and Technology (ICETECH). Coimbatore, India; IEEE: #7275024 [DOI: 10.1109/ICETECH.2015.7275024]
- Peng Y K and Ling Z H. 2022. Decoupled pronunciation and prosody modeling in meta-learning-based multilingual speech synthesis [EB/OL]. [2022-09-14]. <https://arxiv.org/pdf/2209.06789.pdf>
- Powers A, Kiesler S, Fussell S and Torrey C. 2007. Comparing a computer agent with a humanoid robot//Proceedings of 2007 ACM/IEEE International Conference on Human-Robot Interaction. Arlington, USA: ACM: 145-152 [DOI: 10.1145/1228716.1228736]
- Prentice C, Dominique Lopes S and Wang X Q. 2020. The impact of artificial intelligence and employee service quality on customer satisfaction and loyalty. *Journal of Hospitality Marketing and Management*, 29(7): 739-756 [DOI: 10.1080/19368623.2020.1722304]
- Prescott T J, Camilleri D, Martinez-Hernandez U, Damianou A and Lawrence N D. 2019. Memory and mental time travel in humans and social robots. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771): #20180025 [DOI: 10.1098/rstb.2018.0025]
- Price B A, Baecker R M and Small I S. 1993. A principled taxonomy of software visualization. *Journal of Visual Languages and Computing*, 4(3): 211-266 [DOI: 10.1006/jvlc.1993.1015]
- Pripfl J, Körtner T, Batko-Klein D, Hebesberger D, Weninger M, Gisinger C, Frennert S, Efring H, Antona M, Adami I, Weiss A, Bajones M and Vincze M. 2016. Results of a real world trial with a mobile social service robot for older adults//Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Christchurch, New Zealand: IEEE: 497-498 [DOI: 10.1109/HRI.2016.7451824]
- Pu X Y, Kross S, Hofman J M and Goldstein D G. 2021. Datamations: animated explanations of data analysis pipelines//Proceedings of 2021 CHI Conference on Human Factors in Computing Systems. Yokohama, Japan: ACM: #3445063 [DOI: 10.1145/3411764.3445063]
- Qi J, Yang P, Waraich A, Deng Z K, Zhao Y B and Yang Y. 2018. Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: a systematic review. *Journal of Biomedical Informatics*, 87: 138-153 [DOI: 10.1016/j.jbi.2018.09.002]
- Qian Y M, Gong X and Huang H J. 2022. Layer-wise fast adaptation for end-to-end multi-accent speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30: 2842-2853 [DOI: 10.1109/TASLP.2022.3198546]
- Qian Y M and Zhou Z K. 2022. Optimizing data usage for low-resource speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30: 394-403 [DOI: 10.1109/TASLP.2022.3140552]
- Qin L B, Xu X, Che W X and Liu T. 2020a. AGIF: an adaptive graph-interactive framework for joint multiple intent detection and slot filling [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/2004.10087.pdf>
- Qin L B, Xu X, Che W X, Zhang Y and Liu T. 2020b. Dynamic fusion network for multi-domain end-to-end task-oriented dialog//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics: 6344-6354 [DOI: 10.18653/v1/2020.acl-main.565]
- Rabiner L and Juang B H. 1993. *Fundamentals of Speech Recognition*. Englewood Cliffs, USA: Prentice-Hall, Inc.
- Rae I, Takayama L and Mutlu B. 2013. The influence of height in robot-mediated communication//Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Tokyo, Japan: IEEE: 1-8 [DOI: 10.1109/HRI.2013.6483495]
- Rae I, Mutlu B and Takayama L. 2014. Bodies in motion: mobility, presence, and task awareness in telepresence//Proceedings of 2014 SIGCHI Conference on Human Factors in Computing Systems. Toronto, Canada: Association for Computing Machinery: 2153-2162 [DOI: 10.1145/2556288.2557047]
- Rae I, Venolia G, Tang J C and Molnar D. 2015. A framework for understanding and designing telepresence//Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing. Vancouver, Canada: Association for Computing Machinery: 1552-1566 [DOI: 10.1145/2675133.2675141]
- Reynolds D A, Quatieri T F and Dunn R B. 2000. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10(1/3): 19-41 [DOI: 10.1006/dspr.1999.0361]
- Reynolds D A and Rose R C. 1995. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1): 72-83 [DOI: 10.1109/89.365379]
- Roberts J and Arnold D. 2012. Robots, the internet and teaching history in the age of the NBN and the Australian curriculum. *Teaching History*, 46(4): 32-34
- Rodríguez-Guerra D, Sorrosal G, Cabanes I and Calleja C. 2021. Human-robot interaction review: challenges and solutions for modern industrial environments. *IEEE Access*, 9: 108557-108578 [DOI: 10.1109/ACCESS.2021.3099287]
- Rohdin J, Stafylakis T, Silnova A, Zeinali H, Burget L and Plchot O. 2019. Speaker verification using end-to-end adversarial language adaptation//Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. Brighton, UK: IEEE: 6006-6010 [DOI: 10.1109/ICASSP.2019.8683616]
- Ronao C A and Cho S B. 2016. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, 59: 235-244 [DOI: 10.1016/j.eswa.2016.04.032]
- Sadri A, Salim F D, Ren Y L, Shao W, Krumm J C and Mascolo C. 2018. What will you do for the rest of the day? An approach to con-

- tinuous trajectory prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4): #186 [DOI: 10.1145/3287064]
- Sakoe H and Chiba S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1): 43-49 [DOI: 10.1109/TASSP.1978.1163055]
- Salichs M A, Castro-González Á, Salichs E, Fernández-Rodicio E, Maroto-Gómez M, Gamboa-Montero J J, Marques-Villarroya S, Castillo J C, Alonso-Martín F and Malfaz M. 2020. Mini: a new social robot for the elderly. *International Journal of Social Robotics*, 12(6): 1231-1249 [DOI: 10.1007/s12369-020-00687-0]
- Samani H, Saadatian E, Pang N, Polydorou D, Fernando O N N, Nakatsu R and Koh J T K V. 2013. Cultural robotics: the culture of robotics and robotics in culture. *International Journal of Advanced Robotic Systems*, 10(12): #400 [DOI: 10.5772/57260]
- Sano A, Phillips A J, Yu A Z, McHill A W, Taylor S, Jaques N, Czeisler C A, Klerman E B and Picard R W. 2015. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones//*Proceedings of the 12th IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. Cambridge, USA: IEEE: 1-6 [DOI: 10.1109/BSN.2015.7299420]
- Seide F, Li G and Yu D. 2011. Conversational speech transcription using context-dependent deep neural networks//*Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Florence, Italy: [s. n.]: 437-440 [DOI: 10.21437/Interspeech.2011-169]
- Shao W, Nguyen T, Qin K, Youssef M and Salim F D. 2018. BLEDoorGuard: a device-free person identification framework using bluetooth signals for door access. *IEEE Internet of Things Journal*, 5(6): 5227-5239 [DOI: 10.1109/JIOT.2018.2868243]
- Shao W, Salim F D, Nguyen T and Youssef M. 2017. Who opened the room? Device-free person identification using bluetooth signals in door access//*Proceedings of 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. Exeter, UK: IEEE: 68-75 [DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.16]
- Shao Z H, Wu Z Q and Huang M L. 2022. AdvExpander: generating natural language adversarial examples by expanding text. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30: 1184-1196 [DOI: 10.1109/TASLP.2021.3129339]
- Sheridan T B. 2016. Human-robot interaction: status and challenges. *Human Factors*, 58(4): 525-532 [DOI: 10.1177/0018720816644364]
- Shi J, Qin K J and Guo K Y. 2018. Research on automated driving classification methods and related test evaluation techniques. *Auto Industry Research*, (7): 30-37 (石娟, 秦孔建, 郭魁元. 2018. 自动驾驶分级方法及相关测试评价技术研究. *汽车工业研究*, (7): 30-37) [DOI: 10.3969/j.issn.1009-847X.2018.07.006]
- Shomin M, Forlizzi J and Hollis R. 2015. Sit-to-stand assistance with a balancing mobile robot//*Proceedings of 2015 IEEE International Conference on Robotics and Automation (ICRA)*. Seattle, USA: IEEE: 3795-3800 [DOI: 10.1109/ICRA.2015.7139727]
- Shrestha N, Barik T and Parnin C. 2021. Unravel: a fluent code explorer for data wrangling//*Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*. Virtual Event, USA: ACM: 198-207 [DOI: 10.1145/3472749.3474744]
- Siddhant A, Goyal A and Metallinou A. 2019. Unsupervised transfer learning for spoken language understanding in intelligent agents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1): 4959-4966 [DOI: 10.1609/aaai.v33i01.33014959]
- Singer P W. 2009. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. New York, USA: Penguin
- Slade P, Tambe A and Kochenderfer M J. 2021. Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision. *Science Robotics*, 6(59): #eabg6594 [DOI: 10.1126/scirobotics.abg6594]
- Snyder D, Garcia-Romero D, Sell G, Povey D and Khudanpur S. 2018. X-vectors: robust DNN embeddings for speaker recognition//*Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, Canada: IEEE: 5329-5333 [DOI: 10.1109/ICASSP.2018.8461375]
- Song Y P, Liu Z Q, Bi W, Yan R and Zhang M. 2020. Learning to customize model structures for few-shot dialogue generation tasks//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics: 5832-5841 [DOI: 10.18653/v1/2020.acl-main.517]
- Sotelo J, Mehri S, Kumar K, Santos J F, Kastner K, Courville A C and Bengio Y. 2017. Char2Wav: end-to-end speech synthesis//*Proceedings of the 5th International Conference on Learning Representations*. Toulon, France: [s. n.]
- Stahnke J, Dörk M, Müller B and Thom A. 2016. Probing projections: interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE Transactions on Visualization and Computer Graphics*, 22(1): 629-638 [DOI: 10.1109/TVCG.2015.2467717]
- Tanaka F, Isshiki K, Takahashi F, Uekusa M, Sei R and Hayashi K. 2015. Pepper learns together with children: development of an educational application//*Proceedings of the 15th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. Seoul, Korea (South): IEEE: 270-275 [DOI: 10.1109/HUMANOIDS.2015.7363546]
- Tomashenko N and Estève Y. 2018. Evaluation of feature-space speaker adaptation for end-to-end acoustic models//*Proceedings of the 11th International Conference on Language Resources and Evaluation*

- (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA)
- Tu Y Z, Mak M W and Chien J T. 2020. Variational domain adversarial learning with mutual information maximization for speaker verification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28: 2013-2024 [DOI: 10.1109/TASLP. 2020.3004760]
- van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A and Kavukcuoglu K. 2016. WaveNet: a generative model for raw audio [EB/OL]. [2016-09-19]. <https://arxiv.org/pdf/1609.03499.pdf>
- Varela-Aldás J, Guamán J, Paredes B and Chicaiza F A. 2020. Robotic cane for the visually impaired//*Proceedings of the 14th International Conference on Universal Access in Human-Computer Interaction*. Copenhagen, Denmark: Springer: 506-517 [DOI: 10.1007/978-3-030-49282-3_36]
- Variani E, Lei X, McDermott E, Moreno I L and Gonzalez-Dominguez J. 2014. Deep neural networks for small footprint text-dependent speaker verification//*Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing*. Florence, Italy: IEEE: 4052-4056 [DOI: 10.1109/ICASSP.2014.6854363]
- Wainer J, Feil-Seifer D J, Shell D A and Mataric M J. 2007. Embodiment and human-robot interaction: a task-based perspective//*The 16th IEEE International Symposium on Robot and Human Interactive Communication*. Jeju, Korea (South): IEEE: 872-877 [DOI: 10.1109/ROMAN.2007.4415207]
- Wan L, Wang Q, Papir A and Moreno I L. 2018. Generalized end-to-end loss for speaker verification//*Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, Canada: IEEE: 4879-4883 [DOI: 10.1109/ICASSP.2018.8462665]
- Wang B Y, Liang Y, Xu D Z, Wang Z H and Ji J. 2021a. Design on electrohydraulic servo driving system with walking assisting control for lower limb exoskeleton robot. *International Journal of Advanced Robotic Systems*, 18 (1) : #172988142199228 [DOI: 10.1177/1729881421992286]
- Wang L, Yu Z W, Guo B, Ku T and Yi F. 2017a. Moving destination prediction using sparse dataset: a mobility gradient descent approach. *ACM Transactions on Knowledge Discovery from Data*, 11(3) : #37 [DOI: 10.1145/3051128]
- Wang L H, Mohammed A and Onori M. 2014. Remote robotic assembly guided by 3D models linking to a real robot. *CIRP Annals*, 63(1) : 1-4 [DOI: 10.1016/j.cirp.2014.03.013]
- Wang L Y, Zhao J X and Zhang L J. 2021b. NavDog: robotic navigation guide dog via model predictive control and human-robot modeling//*Proceedings of the 36th Annual ACM Symposium on Applied Computing*. Virtual Event Republic of Korea: Association for Computing Machinery: 815-818 [DOI: 10.1145/3412841.3442098]
- Wang M N. 2019. Research on the Effect of Joint Attention Intervention of Autistic Children Based on Social Robot. Hangzhou: Zhejiang University of Technology (王蒙娜. 2019. 基于社交机器人的孤独症儿童共同注意干预效果研究. 杭州: 浙江工业大学) [DOI: 10.27463/d.cnki.gzgyu.2019.000530]
- Wang R, Wang W C, DaSilva A, Huckins J F, Kelley W M, Heather-ton T F and Campbell A T. 2018. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1) : #43 [DOI: 10.1145/3191775]
- Wang R Y, Wang S H, Du S Y, Xiao E D, Yuan W Z and Feng C. 2020a. Real-time soft body 3D proprioception via deep vision-based sensing. *IEEE Robotics and Automation Letters*, 5(2) : 3382-3389 [DOI: 10.1109/LRA.2020.2975709]
- Wang T, Tao J H, Fu R B, Yi J Y, Wen Z Q and Zhong R X. 2020b. Spoken content and voice factorization for few-shot speaker adaptation//*Proceedings of the 21st Annual Conference of the International Speech Communication Association*. Shanghai, China: [s. n.]: 796-800
- Wang W S, Na X X, Cao D P, Gong J W, Xi J Q, Xing Y and Wang F Y. 2020c. Decision-making in driver-automation shared control: a review and perspectives. *IEEE/CAA Journal of Automatica Sinica*, 7(5) : 1289-1307 [DOI: 10.1109/JAS.2020.1003294]
- Wang Y X, Skerry-Ryan R J, Stanton D, Wu Y H, Weiss R J, Jaitly N, Yang Z H, Xiao Y, Chen Z F, Bengio S, Le Q, Agiomyrgi-annakis Y, Clark R and Saurous R A. 2017b. Tacotron: towards end-to-end speech synthesis//*Proceedings of INTERSPEECH 2017*. Stockholm, Sweden: ISCA: 4006-4010 [DOI: 10.21437/Interspeech.2017-1452]
- Wei H B. 2022. Semi-Autonomous Teleoperated Robot System for Power Grid Operations. Guangzhou: Guangdong University of Technology (韦海彬. 2022. 面向电网作业的半自主遥操作机器人系统研究. 广州: 广东工业大学) [DOI: 10.27029/d.cnki.ggdgu.2022.000296]
- Wei K, Zhang Y K, Sun S N, Xie L and Ma L. 2022a. Leveraging acoustic contextual representation by audio-textual cross-modal learning for conversational ASR [EB/OL]. [2022-07-03]. <https://arxiv.org/pdf/2207.01039v1.pdf>
- Wei K, Zhang Y K, Sun S N, Xie L and Ma L. 2022b. Conversational speech recognition by learning conversation-level characteristics//*Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore, Singapore: IEEE: 6752-6756 [DOI: 10.1109/ICASSP43922.2022.9746884]
- Wei Y T, Mei H H, Huang W Q, Wu X Y, Xu M L and Chen W. 2022c. An evolutionary model for operation-driven visualization design. *Journal of Visualization*, 25(1) : 95-110 [DOI: 10.1007/s12650-021-00784-w]
- Wilk R and Johnson M J. 2014. Usability feedback of patients and therapists on a conceptual mobile service robot for inpatient and home-based stroke rehabilitation//*Proceedings of the 5th IEEE RAS/*

- EMBS International Conference on Biomedical Robotics and Biomechatronics. Sao Paulo, Brazil: IEEE: 438-443 [DOI: 10.1109/BIOROB.2014.6913816]
- Witt P L, Wheelless L R and Allen M. 2004. A meta-analytical review of the relationship between teacher immediacy and student learning. *Communication Monographs*, 71 (2) : 184-207 [DOI: 10.1080/036452042000228054]
- Wu C S, Socher R and Xiong C M. 2019a. Global-to-local memory pointer networks for task-oriented dialogue [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/1901.04713.pdf>
- Wu P F, Ling Z H, Liu L J, Jiang Y, Wu H C and Dai L R. 2019b. End-to-end emotional speech synthesis using style tokens and semi-supervised training//*Proceedings of 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. Lanzhou, China: IEEE: 623-627 [DOI: 10.1109/APSIPAASC47483.2019.9023186]
- Wu Z Z and King S. 2016. Investigating gated recurrent networks for speech synthesis//*Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai, China: IEEE: 5140-5144 [DOI: 10.1109/ICASSP.2016.7472657]
- Xiao A X, Tong W Z, Yang L Z, Zeng J, Li Z Y and Sreenath K. 2021. Robotic guide dog: leading a human with leash-guided hybrid physical interaction//*Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE: 11470-11476 [DOI: 10.1109/ICRA48506.2021.9561786]
- Xiao H. 2020. China's new energy vehicles enter a new phase of accelerated development. *Sinopec Monthly*, (11): #75 (萧河. 2020. 中国新能源汽车进入加速发展新阶段. *中国石化*, (11): #75) [DOI: 10.3969/j.issn.1005-457X.2020.11.025]
- Xiong K, Fu S W, Ding G M, Luo Z S, Yu R, Chen W, Bao H J and Wu Y C. 2022. Visualizing the scripts of data wrangling with SOMNUS. *IEEE Transactions on Visualization and Computer Graphics* [DOI: 10.1109/TVCG.2022.3144975]
- Xu T. 2022. Research on the Design of Human-Computer Interface in the Situation of Autonomous Driving Take-Over. Mianyang: Southwest University of Science and Technology (徐韬. 2022. 自动驾驶接管情境下的人机交互界面设计研究. 绵阳: 西南科技大学) [DOI: 10.27415/d.cnki.gxngc.2022.000665]
- Xu Y. 2021. Spatial memory and visual characteristics of drivers in L2 autonomous driving scenarios. Tianjin: Tianjin Normal University (徐杨. 2021. L2自动驾驶情境下驾驶员空间记忆与视觉特征. 天津: 天津师范大学) [DOI: 10.27363/d.cnki.gtsfu.2021.000611]
- Yalçın M A, Elmqvist N and Bederson B B. 2018. Keshif: rapid and expressive tabular data exploration for novices. *IEEE Transactions on Visualization and Computer Graphics*, 24 (8) : 2339-2352 [DOI: 10.1109/TVCG.2017.2723393]
- Yang C Y, Zhou S R, Guo J L C and Kästner C. 2021. Subtle bugs everywhere: generating documentation for data wrangling code//*Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. Melbourne, Australia: IEEE: 304-316 [DOI: 10.1109/ASE51524.2021.9678520]
- Yang L Q, Ting K and Srivastava M B. 2014. Inferring occupancy from opportunistically available sensor data//*Proceedings of 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. Budapest, Hungary: IEEE: 60-68 [DOI: 10.1109/PerCom.2014.6813945]
- Yang Q Y. 2022. Motion Control Strategy and Model Optimization based on Master-Slave Teleoperation Robot. Nanning: Guangxi University (杨启业. 2022. 基于主从遥操作机器人的运动控制策略及模型优化研究. 南宁: 广西大学) [DOI: 10.27034/d.cnki.ggxju.2022.001210]
- Yang X D and Tian Y L. 2017. Super normal vector for human activity recognition with depth cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5) : 1028-1039 [DOI: 10.1109/TPAMI.2016.2565479]
- Yeh C F, Mahadeokar J, Kalgaonkar K, Wang Y Q, Le D, Jain M, Schubert K, Fuegen C and Seltzer M L. 2019. Transformer-transducer: end-to-end speech recognition with self-attention [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/1910.12977.pdf>
- Yeh C K, Chen J S, Yu C Z and Yu D. 2018. Unsupervised speech recognition via segmental empirical output distribution matching [EB/OL]. [2023-01-13]. <https://arxiv.org/pdf/1812.09323.pdf>
- Yu D, Yao K S, Su H, Li G and Seide F. 2013. KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition//*Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, Canada: IEEE: 7893-7897 [DOI: 10.1109/ICASSP.2013.6639201]
- Yu Z W and Wang Z. 2020. Human Behavior Analysis: Sensing and Understanding. Singapore, Singapore: Springer [DOI: 10.1007/978-981-15-2109-6]
- Yuan W, Li Z J and Su C Y. 2021. Multisensor-based navigation and control of a mobile service robot. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51 (4) : 2624-2634 [DOI: 10.1109/TSMC.2019.2916932]
- Zhai X H, Oliver A, Kolesnikov A and Beyer L. 2019. S4L: self-supervised semi-supervised learning//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South) : IEEE: 1476-1485 [DOI: 10.1109/ICCV.2019.00156]
- Zhang B Q, Barbareschi G, Herrera R R, Carlson T and Holloway C. 2022a. Understanding interactions for smart wheelchair navigation in crowds//*Proceedings of 2022 CHI Conference on Human Factors in Computing Systems*. New Orleans, USA: Association for Computing Machinery: #194 [DOI: 10.1145/3491102.3502085]
- Zhang F S, Chang Z X, Niu K, Xiong J, Jin B H, Lyu Q and Zhang D

- Q. 2020a. Exploring LoRa for long-range through-wall sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2): #86 [DOI: 10.1145/3397326]
- Zhang H J and Qian J. 2019. Design of telepresence mobile robot system based on ROS. *Manufacturing Automation*, 41(9): 111-114 (张华健, 钱钧. 2019. 基于ROS的远程呈现移动机器人系统设计. *制造业自动化*, 41(9): 111-114) [DOI: 10.3969/j.issn.1009-0134.2019.09.025]
- Zhang K L, Liu T T, Liu Z, Zhuang Y and Chai Y J. 2020. Multimodal human-computer interactive technology for emotion regulation. *Journal of Image and Graphics*, 25(11): 2451-2464 (张凯乐, 刘婷婷, 刘箴, 庄寅, 柴艳杰. 2020. 面向情绪调节的多模态人机交互技术. *中国图象图形学报*, 25(11): 2451-2464) [DOI: 10.11834/jig.200251]
- Zhang Q, Lu H, Sak H, Tripathi A, McDermott E, Koo S and Kumar S. 2020b. Transformer transducer: a streamable speech recognition model with transformer encoders and RNN-T loss//*Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE: 7829-7833 [DOI: 10.1109/ICASSP40776.2020.9053896]
- Zhang R S, Zheng Y H, Shao J Z, Mao X X, Xi Y D and Huang M L. 2020c. Dialogue distillation: open-domain dialogue augmentation using unpaired data//*Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing*. Virtual Event Association for Computational Linguistics: 3449-3460 [DOI: 10.18653/v1/2020.emnlp-main.277]
- Zhang X, Li W Z, Chen X and Lu S L. 2018. MoodExplorer: towards compound emotion detection via smartphone sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4): #176 [DOI: 10.1145/3161414]
- Zhang X S. 2013. *The Design And Development For Visual Programming In Mobile Devices*. Taiyuan, Shanxi Normal University. (张秀深. 2013. 可视化编程移动学习平台研究与实现. 太原: 山西师范大学)
- Zhang Y, Lyu Z Q, Wu H B, Zhang S S, Hu P F, Wu Z Y, Lee H Y and Meng H L. 2022b. MFA-conformer: multi-scale feature aggregation conformer for automatic speaker verification [EB/OL]. [2022-11-11]. <https://arxiv.org/pdf/2203.15249.pdf>
- Zhang Y, Li Z Y, Guo H L, Wang L Y, Chen Q H, Jiang W J, Fan M M, Zhou G Y and Gong J T. 2023. "I am the follower, also the boss": exploring different levels of autonomy and machine forms of guiding robots for the visually impaired//*Proceedings of 2023 CHI Conference on Human Factors in Computing Systems*. Hamburg, Germany. ACM: 1-22 [DOI: 10.1145/3544548.3580884]
- Zhao Y G, Li M R, Li B H and Han B. 2010. A method and implementation of program visualization based on speculative multithreading technology. *Journal of Xi'an University of Posts and Telecommunications*, 15(5): 69-74 (赵永刚, 李美蓉, 李保红, 韩博. 2010. 基于推测多线程技术的程序可视化方法与实现. *西安邮电学院学报*, 15(5): 69-74) [DOI: 10.13682/j.issn.2095-6533.2010.05.028]
- Zheng Y, Li Q N, Chen Y K, Xie X and Ma W Y. 2008. Understanding mobility based on GPS data//*Proceedings of the 10th International Conference on Ubiquitous Computing*. Seoul, Korea (South): ACM: 312-321 [DOI: 10.1145/1409635.1409677]
- Zhong Z, Lei M Y, Cao D L, Fan J P and Li S Z. 2017. Class-specific object proposals re-ranking for object detection in automatic driving. *Neurocomputing*, 242: 187-194 [DOI: 10.1016/j.neucom.2017.02.068]

作者简介

陶建华,男,教授,主要研究方向为语音处理、认知推理、数据内容分析和智能交互。E-mail: jhtao@tsinghua.edu.cn

梁山,通信作者,男,博士,主要研究方向为语音信息处理、语音伪造检测、传感器阵列语音信息处理。

E-mail: sliang@nlpr.ia.ac.cn

龚江涛,女,助理研究员,助理教授,博士,主要研究方向为人—机器人交互、类人认知计算、虚实融合和无障碍计算。

E-mail: gongjiangtao@air.tsinghua.edu.cn

高楠,女,博士后,主要研究方向为普适计算、心理感知和人机交互。E-mail: nangao@tsinghua.edu.cn

傅四维,男,副研究员,主要研究方向为人机交互、信息可视化和可视分析。E-mail: siwei.fu@zhejianglab.com

喻纯,男,副教授,主要研究方向为智能人机交互。

E-mail: chunyu@tsinghua.edu.cn