

“智能人机交互前沿技术丛书”序

随着智能化时代的到来,人机交互(human-computer interaction, HCI)的发展将进入新的阶段,并将深刻影响并改变人类社会的运行模式和未来发展路径。作为人工智能与人机交互深度融合的产物,智能人机交互正逐步成为连接人类智慧与机器智能的桥梁,对全面提升人类理解世界、预测世界、塑造世界和调和世界的的能力起着至关重要的作用。

智能人机交互技术的核心价值在于其能够实现人与机器之间更加自然、高效、协同的交互方式。这不仅要求机器能够准确理解人类的意图与行为,还须具备适应复杂环境、进行灵活反馈的能力。随着人工智能、大数据、云计算、扩展现实等技术的快速发展,智能人机交互正向着更加智能化、个性化的方向迈进,为人们带来前所未有的“如同手足”的自然操控体验、“身临其境”的场景感知,以及“心有灵犀”的高效互动。

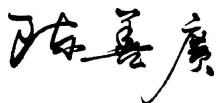
鉴于此,世界各主要国家都高度重视人机交互技术的发展。2024年,美国发布《关键与新兴技术国家战略》,将人机交互技术列为18项关键与新兴技术之一。人机交互技术一直以来也是我国重大战略需求前沿,国家自然科学基金委员会信息科学部将其列为优先发展领域。

近年来,我国智能人机交互技术研究与应用取得了可喜的发展与长足的进步,涌现出一批具有世界水平的理论研究成果,突破了一系列关键技术,开展了效益显著的应用探索,也培育了大批富有创新意识和创新能力的人才。

“智能人机交互前沿技术丛书”是军事科学院国防科技创新研究院联合多家国内的顶级科研院所在长期跟踪我国科技发展前沿,广泛征求专家意见的基础上,经反复论证、精心选题后组织出版的。该丛书聚焦于人机交互领域的最新研究成果与技术突破,旨在通过多维度、多层次的视角,全面展现智能人机交互的理论基础、关键技术、创新应用以及未来趋势,既有对人机交互模式、界面设计、系统架构等核心问题的深入探讨,也有对人工智能技术在人机交互中的应用创新,以期突出我国在该重点前沿领域取得的基础性、原创性和变革性成果。

我相信在大家的共同努力下,该丛书将成为人机交互技术领域的重要参考书目,为推动该领域的发展做出积极贡献。

同时,衷心希望该丛书的出版能够激发更多人对这一领域的兴趣与热情,吸引更多有志之士加入智能人机交互的研究与实践中来,携手共进,探索未知,勇于创新,共同开创智能人机交互技术更加美好的未来!

A stylized handwritten signature in black ink, reading '陈勇广' (Chen Yonggu).

国际宇航科学院院士

人因工程全国重点实验室主任

中国载人航天工程原副总设计师

前言

人机交互(human-computer interaction, HCI)的历史可以追溯到 20 世纪 60 年代,随着计算机技术的不断进步和普及,人机交互已经成为人们日常生活和工作中不可或缺的一部分。无论是个人计算设备、智能手机、家庭娱乐系统,还是在医疗、教育、交通等领域,人机交互都扮演着重要角色。然而,传统的人机交互仍然存在一些局限性。虽然现有的界面设计和交互方式已经相当成熟,但其仍然无法完全理解人类的语言、情感和意图。通过计算机手段辅以智能算法,使人机交互更加自然、舒适,这就是智能人机交互的研究意义。智能人机交互赋予计算机和智能设备更强大的感知能力和智能化处理能力,使其能够更好地理解和响应人类的需求和意图。通过手势交互、眼动交互、语音交互、脑机情绪交互、跨模态交互等技术,智能人机交互可以实现更加自然、智能、个性化的交流和操作体验。

当今正处于数字化和智能化时代,这是一个与计算机和智能设备紧密互动的世界,人机交互技术的质量对人们的日常生活、工作和娱乐产生了深远影响。但是,在智能人机交互的研究中尚缺乏能够兼顾智能人机交互领域广度和深度的著作。作者研究团队一直以来致力于智能人机交互技术的研究,在智能人机交互的范式、算法、实验和工程实践等方面积累了丰富的经验。着眼于智能人机交互理论和实践的结合,同时注重新兴概念,撰写了本书,希望为从事相关工作的研究人员和工程开发人员提供帮助。

本书注重学术研究和实际应用的结合,旨在为读者深入了解智能人机交互技术的最新进展和未来发展方向提供帮助。本书不仅介绍各种技术的原理和应用场景,还深入探讨背后的理论基础和研究思想,使读者能够更好地理解技术的本质和潜力。

本书共 8 章。第 1 章为本书奠定基调,介绍交互设计、人机交互、人因工程以及智能人机交互的概念,最后引出本书智能人机交互的五大技术:手势、眼动、语音、脑机情绪、跨模态。该章适合从事智能人机交互的初级研究人员,也适合熟悉基础,希望深入了解人机交互概念的高级研究人员。

第 2 章至第 6 章,对智能人机交互的五大技术(手势、眼动、语音、脑机情绪以及跨模态)进行全面介绍。第 2 章全面介绍手势交互的三个方面,包括基于视觉传感器的手势交互、基于穿戴式传感器的手势交互和基于多模态传感器的手势交互。第 3 章全面介绍眼动视线估计的四种方法:基于姿态偏移量校正的单被试视线估计方法、基于 AdaBN 的跨被试视线估计方法、基于差分的双眼不对称模型和基于

头眼协同的视线追踪技术。第4章全面介绍语音识别的四种方法:基于音频信号的语音识别方法、基于视觉图像的语音识别方法、基于表面肌电的语音识别方法和基于多模态融合的语音识别方法。第5章重点关注脑机情绪中的负面情绪。第6章重点关注“人→机”跨模态语义共指消解、“机→人”图像语义文本生成,以及“人↔机”跨模态智能语义匹配等关键技术。这五章是智能人机交互技术的核心,对其中某一特定主题感兴趣的读者可以从相关的章节中获得有用的信息。

第7章介绍混合现实技术,这是一种使真实世界和虚拟物体在同一视觉空间中显示和交互的计算机技术,融合人机交互、传统现实及人的认知等多个领域,是一门面向应用的交叉学科。

为了便于读者阅读,本书提供部分彩图的电子文件,读者可自行扫描封底二维码查阅。

第8章重点关注可穿戴智能交互系统集成方案,包括智能头盔、智能眼镜、智能手套、智能戒指。此外,介绍两个典型应用系统开发:多人电子沙盘应用和多机交互应用。该章将第2章至第6章提出的技术进行融合,为读者从系统层次更好地了解智能人机交互技术提供启发。

为便于读者阅读,本书提供了部分彩图的电子文件,读者可自行扫描封底二维码查阅。

本书由闫野、印二威、谢良、刘小龙、施忠臣、倪群平共同撰写。课题组成员张皓洋、白晓伟、贺浩、郑昌艳、罗治国、赵煜、马闯等参与了本书相关的理论研究、系统设计、安装调试和测试应用等工作。课题组成员詹涵灵,王梓衡、吴竞寒、常惠琴、胡小荣、刘闪闪、冯登放等承担了资料整理、版面编排、文字校对、图表绘制、文献核查等辅助工作。另外特别感谢杨双老师、任鹏飞老师对本书做出的指导工作。

在本书撰写过程中,汇集了各领域专家和研究者的知识与经验,努力提供准确、全面的信息。希望本书能够激发读者对智能人机交互技术的兴趣,并能够成为读者的指南,帮助其理解、应用和推动智能人机交互技术的发展,为数字化未来开辟更广阔的可能性。

由于作者水平有限,书中不足之处在所难免,恳请读者批评指正。

目 录

“智能人机交互前沿技术丛书”序

前言

第 1 章 绪论	1
1.1 交互设计	1
1.2 人机交互	3
1.3 人因工程	6
1.4 智能人机交互	8
参考文献	11
第 2 章 手势交互	13
2.1 手势交互概述	13
2.1.1 手势交互基本概念	13
2.1.2 基于穿戴式传感器的手势交互方法	15
2.1.3 基于视觉传感器的手势交互方法	16
2.1.4 基于多模态传感器的手势交互方法	18
2.2 基于视觉传感器的手势交互	19
2.2.1 基于深度图像的手部姿态估计研究	19
2.2.2 基于 RGB 图像的手部姿态估计研究	24
2.2.3 基于 RGB-D 图像的手部姿态估计研究	33
2.3 基于穿戴式传感器的手势交互	37
2.3.1 基于 IMU 的手势交互	37
2.3.2 基于 sEMG 信号的手势交互	56
2.4 基于多模态传感器的手势交互	66
2.4.1 多模态手势数据	66
2.4.2 基于多模态传感器融合的手势识别方法	68
参考文献	89
第 3 章 眼动交互	95
3.1 眼动交互概述	95
3.1.1 早期眼动追踪方法	95
3.1.2 现代眼动追踪方法	98
3.1.3 主流的眼动追踪设备	100

3.2	大型近眼视线估计数据集	100
3.2.1	数据集采集设备	101
3.2.2	数据采集范式	102
3.3	基于姿态偏移量校正的单被试视线估计方法	104
3.3.1	基础模型	104
3.3.2	基于姿态偏移量的模型改进	107
3.3.3	模型表现	110
3.4	基于 AdaBN 的跨被试视线估计方法	113
3.4.1	AdaBN 介绍	114
3.4.2	基于 AdaBN 的模型改进	116
3.4.3	离线实验	117
3.5	基于差分的双眼不对称模型	122
3.5.1	基础模型	122
3.5.2	不对称双眼模型	122
3.5.3	差分学习	124
3.5.4	实验对比	126
3.6	基于头眼协同的视线追踪技术	126
3.6.1	头眼协同运动分析	127
3.6.2	头眼协同视线追踪网络	130
3.6.3	模型训练与部署	132
3.6.4	实验结果与评估	133
	参考文献	136
第 4 章	语音交互	139
4.1	语音交互概述	139
4.1.1	基于音频信号的语音识别	141
4.1.2	基于视觉图像的无声语音识别	142
4.1.3	基于 sEMG 信号的无声语音识别	143
4.1.4	多模态语音识别	145
4.2	基于音频信号的语音识别方法	146
4.2.1	音频信号预处理	146
4.2.2	声学建模方法	149
4.2.3	语言模型	155
4.2.4	基于音频信号的自动语音识别模型	157
4.3	基于视觉图像的语音识别方法	160
4.3.1	唇部图像空域特征提取	160

4.3.2	唇部图像时域特征提取	165
4.3.3	唇语识别方法设计	171
4.4	基于 sEMG 信号的语音识别方法	178
4.4.1	单词级/指令词识别	178
4.4.2	连续无声语音识别	186
4.5	基于多模态融合的语音识别方法	190
4.5.1	视觉肌电信息融合的多模态无声语音识别	190
4.5.2	视听融合的多模态语音识别	197
4.5.3	声光电多模态融合的语音识别	201
	参考文献	210
第 5 章	脑机情绪交互	215
5.1	脑机情绪交互概述	215
5.1.1	脑机接口	215
5.1.2	脑机接口三大范式	216
5.1.3	情感脑机接口	219
5.2	多维负性情绪素材库和数据集	221
5.2.1	情绪模型	221
5.2.2	情绪诱发	223
5.2.3	情绪诱发实验被试	225
5.2.4	自我评估问卷和统计分析	226
5.2.5	情绪诱发实验设计	227
5.3	多维负性情绪识别方法	229
5.3.1	表征负性情绪变化的脑电图及其他生理信号	229
5.3.2	预处理方法	232
5.3.3	特征提取方法	234
5.3.4	负性情绪识别相关工作	237
5.3.5	问题及展望	246
5.4	基于脑机接口的广域目标识别技术	247
5.4.1	基于快速序列视觉呈现的脑机接口系统	248
5.4.2	空间选择性注意相关的 ERP 成分	250
5.4.3	单因素空间编码的广域目标识别	251
5.4.4	多因素空间编码的广域目标识别	255
5.5	运动任务中的数据增强与频率特征提取	257
5.5.1	相关方法与数据集构建	257
5.5.2	基于脑区重组的 EEG 数据增强方法	261

5.5.3 基于张量的脑机接口频率特征组合方法	269
参考文献	279
第6章 跨模态交互	289
6.1 跨模态交互概述	289
6.2 “人→机”跨模态语义共指消解	290
6.2.1 图像+文本共指消解	290
6.2.2 语音+眼动+图像共指消解	295
6.3 “机→人”图像语义文本生成	301
6.3.1 图像文本生成	302
6.3.2 空间语义学	304
6.3.3 数据集构建	306
6.3.4 实验结果	307
6.4 “人↔机”跨模态智能语义匹配	309
6.4.1 指令级跨模态匹配	311
6.4.2 子指令级跨模态匹配	318
6.4.3 实体级跨模态匹配	320
6.4.4 指令语义解析任务	321
参考文献	324
第7章 混合现实技术	328
7.1 系统组成	328
7.1.1 HoloLens 系列	330
7.1.2 Magic Leap One	332
7.1.3 Nreal 系列	333
7.2 光学显示技术	334
7.2.1 微显示器	334
7.2.2 典型光学显示结构分析	335
7.2.3 光透式 AR 设备标定方案	343
7.3 定位定向技术	348
7.3.1 坐标一致性原理	348
7.3.2 SLAM 中的信息融合技术	358
7.3.3 无标记动作捕捉结合的 IMU 定位技术	363
7.3.4 三维物体刚体位姿估计	366
7.3.5 深度学习定位定向技术	372
7.4 应用案例	373
7.4.1 民生应用	374

7.4.2 军事应用	378
参考文献	383
第 8 章 智能人机交互系统的开发及应用	389
8.1 移动端人工智能算法工具与部署技术栈	389
8.1.1 基于 TensorFlow 的移动端算法部署	390
8.1.2 基于 PyTorch 的移动端算法部署	391
8.1.3 基于高通芯片的 SNPE 框架部署	392
8.1.4 基于 NCNN 加速框架的算法部署及优化	393
8.1.5 基于 MediaPipe 的测试及移植	395
8.2 可穿戴智能交互系统集成方案	396
8.2.1 智能头盔	399
8.2.2 智能眼镜	404
8.2.3 智能手套	410
8.2.4 智能戒指	416
8.3 典型应用系统开发	420
8.3.1 多人电子沙盘应用	420
8.3.2 多机交互应用	428
第 9 章 结束语	436

第 1 章 绪 论

“交互”是一个古老的字眼,《京氏易传》是西汉京房所撰写的经书注释,其中就提到“震分阴阳,交互用事”。“交互”的意思是替换、交换。扩展来讲,交互首先要存在于两个或两个以上对象之间,这里的对象可以指人,也可以指物;其次对象之间要有相互作用,这种相互作用应该是有章可循的、能够形成相互响应和因果反馈的。总的来说,“交互”泛指人与人、人与物、物与物之间的相互作用和影响。

工程技术视角的“交互”技术更多地出现在工业革命之后。第一次工业革命始于 18 世纪 60 年代的英国,随着蒸汽机的普及运用,社会进入了“蒸汽时代”,出现了纺织机、蒸汽机车等一大批复杂机器。19 世纪 70 年代,在欧美主要国家几乎同时兴起了第二次工业革命,世界进入了“电气时代”。电动机的发明和电力的普遍应用再次改变了世界,电灯、电话机、电报机、汽车等重大发明及其背后的基础设施进一步使机器设备深入社会的各个方面。20 世纪中叶,人类开始了第三次工业革命。电子计算机、空间技术和生物工程极大地改变了世界面貌,特别是电子技术及计算机技术,从电子管、晶体管到集成电路,从传统的专用计算机发展到通用计算机、家用台式机、笔记本电脑,再到超级计算机、手机、各种智能家电,已经渗透到人类生活的各个角落。当今时代,机器已经成为人类生活中不可分割的组成部分,如何使人与机器更好地沟通,成为迫切需要解决的问题。

1.1 交互设计

交互设计(interaction design, ID)的初衷是使人和机器能够有一个更加有效的交流手段。对于简单的机器,交互设计相对容易,往往通过一个好用的按钮、一条简单的提示就能实现,对于复杂机器,交互设计则要困难得多。例如,一套涉及多个复杂操作步骤的电气系统是这样设计的:为了避免工人在操作过程中使用错误的操作顺序或遗漏必要的操作环节,工人要拿专门的钥匙解锁操作 1 的面板,只有正确完成了操作面板 1 上的操作才能得到解锁操作面板 2 的钥匙,以此类推,直到十几个操作面板都被正确解锁才能完成全部操作流程。整个过程非常烦琐,但确实有效解决了误操作的问题,是典型的牺牲时间效率换取避免由操作错误引发重大损失的方法。进入计算机时代,人们与机器打交道的复杂性越发凸显,人们更期望在保证正确性的同时使用更加简洁的方式与机器实现双向交流,这就是交互设计需要解决的问题,“以人为核心”是交互设计的基本出发点。

交互设计完全继承了工业设计以用户为中心的思想。交互设计思想强调把机器的工作过程看作一个事件,设计者要重点考察机器在事件中的行为和过程,从机器的流程图、状态转换图和故事板等出发,从多个剖面研究机器可能的行为方式,再从人的使用角度考虑如何构建一个适合自己的人机交互界面。

交互设计就是定义一个人工制品在特定场景下反应方式的相关界面。交互设计作为一门关注交互体验的学科诞生于 20 世纪 80 年代,由比尔·莫格里奇提出,一开始使用的是“软面(soft face)”这一说法,后来更名为“interaction design”,即交互设计。

最初,交互设计这一概念更多的是以界面设计为中心,如网站设计和图形设计,目的是使用户更加容易、毫无察觉地掉入设计者精心设计的“陷阱”,如关注某些特定信息、进入某些特定页面或更长时间地把用户留在网站。现在,这一工作仍然是交互设计的重要部分。例如,某知名视频社交网站就针对手机界面视频的分割方式(包括边框宽度、视频索引大小等)做了专门优化,只是为了让用户在应用中感觉更舒服,多停留一些时间。这只是交互设计中很小的一部分,许多类似的工作在大数据分析的加持下正在悄无声息地进行着,大部分的工作对用户而言几乎是完全透明的。目前,交互设计已经成长为一个独立的领域,交互设计师远非仅负责文字和图像,而是负责创建在机器上的所有交互元素,即用户可能会触摸、点击或者输入的所有内容。简而言之,交互设计面向的是用户使用产品过程中的所有交互环节。

最早的互动技术设计项目是麻省理工学院(Massachusetts Institute of Technology, MIT)的 Cooper 在 1975 年创立的可视语言工作坊,以及由 Elton 于 1979 年在纽约大学创立,后来由 Burns 领导的互动电信项目^[1]。1994 年,卡内基梅隆大学成立了第一个正式命名为“交互设计”的学术方向,并授予交互设计专业学生硕士学位。该类项目最初主要侧重于计算机屏幕界面,随着互联网技术的发展,逐渐拓展到人群与组织、文化、服务、系统的互动。

美国有数十所高校开设了交互设计专业,只是所使用的名称有所差异,如互动媒体艺术、数字交互设计、用户体验设计、信息设计等。美国交互设计专业面向硕士研究生居多,主要是因为硕士研究生项目的培养模式更能满足交互设计专业对多样性人才的需求,可以将不同学科背景及行业经验的学生会聚到一起,以项目为主导,在多知识结构的群体下相互合作、启发学习。

交互设计作为一门学科,特别关注所设计交互系统的可用性,它是对系统可用程度的总体评价,是从用户角度衡量交互产品是否有效、易学、安全、高效、好记、少错的关键性指标。同时,交互设计还要考虑使用者的期望和体验。可用,是产品应该做到的,而体验则可以让人们获得惊喜。认知心理学为交互设计提供了基础设计原则,包括心智模型(mental model)、感知/现实映射(mapping)原理、比喻

(metaphor)以及可操作暗示(affordance)。交互设计的原则包括但不限于以下方面。

(1)可视性。好的功能可视性能够让用户更容易发现和了解如何与产品进行交流。

(2)反馈性。对与活动相关的信息给予及时、必要的响应,以引导用户继续下一步操作。

(3)容错性。当用户误操作时,在特定时刻采取必要的措施,并给予用户提示。

(4)映射性。准确表达引导控制意图及其交互结果之间的对应关系。

(5)一致性。保持同一交互系统同一功能交互界面的表达范式不变。

(6)启发性。用易于理解的表达方式给予用户必要的提示。

总的来说,交互设计是人类工程学、心理学、认知科学、信息学、工程学、计算机科学、软件工程学、社会学、人类学、语言学、美学等众多学科的交叉融合,它以更倾向于人文社科的方式构建人与机器之间的沟通桥梁。

1.2 人 机 交 互

人机交互(human-computer interaction, HCI)主要研究人的认知模型和信息处理过程机器与人的交互行为之间的关系,研究如何依据用户的任务和活动进行交互式计算系统的设计、实现和评估。

相较于现代意义的交互设计,人机交互的概念似乎出现得更早。1959年,美国学者 Shackle^[2]从人类在操纵计算机时如何能减轻疲劳程度出发,撰写了被认为是人机交互的第一篇论文。这是一篇关于如何设计计算机控制台的人机工程学(man-machine engineering, MME)论文。1960年,人机紧密共栖(human-computer close symbiosis)的概念被提出,被视为人机界面学的启蒙。1969年,在英国剑桥大学召开了第一次人机系统国际大会,同年,第一份专业杂志《国际人机研究》创刊。20世纪70年代初,全球首批专注于人机交互设计的先驱性研究机构相继诞生——英国拉夫堡大学的人文科学与先进技术研究所和美国帕罗奥多研究中心分别成为该领域研究的开拓者。从20世纪70年代后期开始,随着个人计算机的普及,人们对人机交互的需求越来越强烈。当时,人机交互往往称为人机界面(human-computer interface, HCI)。到了20世纪80年代,人机交互学科逐渐形成了自身的理论体系和实践架构。在理论体系方面,从人机工程学独立出来,更加强调认知心理学、行为学和社会学的某些人文科学的理论指导;在实践架构方面,从人机界面拓延开来,强调计算机对于人的反馈的交互作用。人机界面(HCI)中的I也由界面/接口(interface)变成了交互(interaction)。尽管人机交互最初起源于计算机科学,但迅速扩展到可视化、信息系统、协同系统、系统开发过程、设计等领域。

从最初关注个人与通用用户扩展到关注社会和组织,从桌面办公应用扩展到游戏、学习、教育、商业、医疗及社区系统,从早期的图形用户界面扩展到更多的交互行为、技术和设备中。

人机交互与交互设计两个学科专业有所区别。从学科属性角度,人机交互在计算机科学领域,强调预先设计好的人与计算机之间的互动关系如何实现,是用户通过输入、处理、输出的循环过程与计算机系统沟通的一种机制;交互设计隶属于设计学科,其信息传递的对象更为广泛,包括人、产品、环境、系统、组织、文化等,通过研究信息传递的方式建立人与外部世界的联系。从研究与实践角度,人机交互从业者往往更注重技术或交互方法模型及其相关软硬件,并且参与科学研究,更多时候是以科学家和工程师的面貌示人;而交互设计师是指引人机交互的思想家和设计者,他们将理论应用到特定的问题、使用环境与目标对象中,提出总体的交互解决框架,更多的是以设计师或艺术家的面貌示人。

以卡内基梅隆大学为例,人机交互与交互设计分别隶属于计算机科学学院的人机交互专业(major of human-computer interaction, MHCI)与设计学院的交互设计专业(major of design, MDes)。MDes 主要研究设计服务与社会创新,通过将产品和信息技术相结合来解决实际的社会问题;而 MHCI 的课程除了包含交互设计的基本理论(如设计思维、流程、方法等)外,更强调技术实现方法的研究,如传感器、控制器、编程、算法分析等。显而易见,人机交互与交互设计具有交叉重合关系,两者都需要在多学科环境中接受教育,包括心理学、商学、人文学、社会科学等。毫无疑问,计算机科学在其中占据着至关重要的地位,因为再好的想法最终都要使用计算机去实现。

可以说,人机交互就是从人类适应计算机过渡到计算机不断地适应人类的发展史。人机交互的发展大致经历了早期手工作业阶段、作业控制语言及交互命令语言阶段、图形用户界面(graphical user interface, GUI)阶段、网络用户界面阶段,直到如今的多通道、多媒体的智能人机交互阶段。

(1)早期手工作业阶段。当时交互的特点是由设计者本人(或本部门同事)来使用计算机,采用手工操作或依赖机器(如输入打孔纸带)的方法去适应当时的计算机。

(2)作业控制语言及交互命令语言阶段。这一阶段的特点是计算机的主要使用者是程序员,他们采用批处理作业语言或交互命令语言与计算机打交道,通过记忆命令,使用键盘、电子管屏幕与计算机交互,并使用非图形化的操作系统(如微软磁盘操作系统(microsoft disk operating system, MS-DOS)),运行专门的编译软件调试程序、了解计算机的执行情况。

(3)图形用户界面阶段。图形用户界面的主要特点是“所见即所得”,使不懂计算机程序的普通用户也可以熟练地使用计算机。图形用户界面的出现使信息产业

得到空前发展。

(4)网络用户界面阶段。以超文本标记语言(hyper text markup language, HTML)及超文本传输协议(hyper text transfer protocol, HTTP)为主要基础的网络浏览器是网络用户界面的代表,由它形成的万维网已经成为当今计算机网络的支柱。这类人机交互技术包括搜索引擎、网络加速、多媒体动画、聊天工具等。

(5)多通道、多媒体的智能人机交互阶段。以虚拟现实为代表的计算机系统的拟人化和以手持电脑、智能手机为代表的计算机系统的微型化、随身化、嵌入化,是当前计算机系统的两个重要发展趋势。目前,虚拟现实领域正在展开一场三维信息技术革命,虚拟现实的交互方式进行了革命性的升级:以鼠标、键盘、手柄为控制器的按键式二维交互升级为以用户自己的头、双手、眼睛为控制器的追踪式三维交互,但是目前虚拟现实中的交互问题影响了虚拟现实的发展。目前,虚拟现实人机交互主要分为五个部分:文字输入^[3]、导航^[4,5]、操纵^[6]、选择^[7]、系统控制。在万物互联的时代,形形色色的计算机使用场景已经渗透到人类生活的每个角落,人类渴望更加自然、高效的人机交互方式。这些交互方式包括但不限于语音、手势、眼动、肌电,甚至脑电波。对各种交互手段的综合运用构成了多通道、多媒体的智能人机交互。虚拟现实人机交互需要使用专门的硬件,如头戴式显示器、运动传感器和触觉反馈设备等,以模拟出一种存在感和沉浸感,并让用户能够以自然和直观的方式与虚拟对象和虚拟环境进行交互。在虚拟现实人机交互中,计算机系统必须能够准确地跟踪用户的动作,并实时响应其输入,以保证身临其境的体验。这就需要先进的计算机视觉、运动跟踪和模拟算法与技术。

当前,智能人机交互技术领域的主要研究范畴包括如下方面。

(1)人机交互界面表示模型与设计方法。人机交互界面设计的好坏直接影响到计算机应用系统的成败。友好型人机交互界面的开发离不开好的交互模型与设计方法。

(2)可用性分析与评估。可用性关系到人机交互能否达到用户期待的目标,以及实现这一目标的效率与便捷性。人机交互系统的可用性分析与评估的研究主要涉及支持可用性的设计原则和可用性的评估方法等。

(3)多通道交互技术。在多通道交互中,用户可以综合使用语音、手势、眼动、肌电等自然的交互方式与计算机进行通信。多通道交互主要研究多通道交互界面的表示模型、评估方法以及多通道信息的融合等。其中,多通道信息的融合是重点和难点。

(4)认知与智能用户界面。智能用户界面的最终目标是使人机交互与人人交互一样自然便捷。上下文感知、眼动跟踪、手势识别、三维输入、语音识别、表情识别、手写识别、自然语言理解等都是认知与智能用户界面需要解决的问题。

(5)群件。群件主要涉及个人或群组间的信息传递、群组中的信息共享、业务

过程自动化与协调,以及人和过程之间的交互活动等。相关研究主要包括群件系统体系结构、共享信息方式、决策支持工具、应用共享以及同步方法等。

(6)Web 互动。Web 互动重点研究 Web 界面的信息交互模型和结构、Web 互动设计工具等。

(7)移动界面设计。移动界面设计针对移动设备的便携性、位置不固定性、计算能力有限性,以及无线网络的低带宽高延迟等诸多限制,应对移动计算、无处不在计算的挑战,研究移动界面的设计方法、导航技术,以及移动界面的实现技术和开发工具。

1.3 人因工程

作为新兴交叉学科,目前人因工程还没有统一、确切的定义,但其在航空航天、核工业等关键领域已经获得了普遍关注。与人因工程最接近的学术领域当属人类工效学,国际人类工效学学会在 2000 年将人因工程定义为“研究人在某种工作环境中的解剖学、生理学和心理学等方面的各种因素;研究人和机器及环境的相互作用;研究在工作中、生活中和休息时怎样统一考虑工作效率、人的健康、安全和舒适等问题的学科”。朱祖祥在《人类工效学》^[8]一书中给出的定义是“一门以心理学、生理学、解剖学、人体测量学等学科为基础,研究如何使人-机-环境系统的设计符合人的身体结构和生理心理特点,以实现人、机、环境之间的最佳匹配,使处于不同条件下的人能有效地、安全地、健康和舒适地工作与生活的科学”。类似的称谓还有人机工程学或人机学,这是中国对“Ergonomics”一词最早的译法。在一些西方国家还经常使用人因工程学(human factors engineering, HFE)的称谓,其最常见的应用场景出现在工业、一般生活领域或生活用品设计中。目前,国内更多地继承了人因工程这一称谓。人因工程的重点在于考察人、机、环境三者之间的关系,并试图将这三者进行有效融合,并强调人的因素在其中居主导地位。

陈善广等^[9]对人因工程这一交叉学科在中国的发展历程进行了回顾和总结,指出人因工程的发展不仅着眼于系统绩效的提高,而是把消除隐患、确保安全置于首位,同时把满足人的多层次需求与系统功能及性能有机统一起来。系统设计缺乏人因工程考虑,将会导致用户不满意,甚至造成人员身体疾患或损伤,导致系统性能下降、效率低下,或者使人发生失误,导致事故。人因工程考虑不周常常造成系统研发周期延长,研发费用增高。

产品研制过程中考虑人因工程不仅不会增加额外成本,反而会提高研制效率。数据统计与分析表明,如果在产品设计早期就考虑人因工程,那么费用仅占总投入的 2%,而如果在产品交付使用发生问题时再引入人因工程进行改进,费用可占总投入的 20%以上。多年来,基于人因工程的人-系统整合(human system integration, HSI)

方法得到美国国防部(Department of Defense, DoD)、核电部门及美国国家航空航天局(National Aeronautics and Space Administration, NASA)等的高度重视和推广应用。实践证明, HSI 在提升系统性能的同时大大节省了系统研发成本, 所获收益包括节省人力资源、降低操作人员的技能需求、缩短训练时间、提高保障和维护效率、降低因人机适配问题造成的安全风险等。

人机关系本质上反映的是人人关系, 因为机是人的创造物, 必然会打上人的烙印; 处理好人机关系就必须处理好设计者与使用者之间的关系, 它是对立统一的。在实践中, 一方面, 设计者常错把自己当作使用者而不考虑真实用户的体验需求; 另一方面, 使用者则期望系统/产品完全了解并满足个人需求, 而不能理解设计者的苦心和工程上的约束, 因此抱怨系统/产品不好用。人因工程为系统/产品的设计者和使用者之间架起了一座科学合理的沟通桥梁。人因工程追求的高安全性、高效率、高满意度、人机融合、缩短周期、降低成本等目标, 无论是对于决策管理者、工程师还是对于最终用户甚至大众都具有重要的意义。

人因工程的四大显著特点包括如下方面^[10]。

(1) 强调以人为中心(human-centered philosophy)。人因工程聚焦一切有人参与的系统、产品或过程, 人是其中的核心; 主要研究人与其他要素的交互规律, 人是设计的出发点和落脚点; 设计必须充分认识并考虑人的特性(生理特点、心理特点、能力与局限等), 充分发挥人的积极作用。

(2) 遵循系统工程方法(system engineering approach)。一方面, 人与其他系统要素交互构成整体, 这里的其他系统要素是指系统中所有的人造物(如工作场所、产品、工具、技术过程、服务、软件、人工环境、任务、组织设计等)和其他人; 另一方面, 人具有不同方面的特性(生理、心理和社会)和不同层面的属性(从操作人员的个体层面到群体组织甚至民族、地区和国家的宏观层面)。人与系统所处的环境也包含物理、社会、信息等不同方面。这些典型的系统特征要求进行人因工程设计分析时必须遵循系统思维和系统工程方法。

(3) 设计驱动(design driven)。人因工程本质上是面向设计、面向系统实现的应用学科, 涵盖策划、设计、实现、评估、维护、再设计和持续改进等阶段。其中, 设计最为关键, 因为 2/3 以上的故障均可追溯到设计源头。人因工程强调与系统研制相关的方面均应参与到策划、设计和研发中, 且人因工程专家应发挥广泛而独特的作用, 例如, 可以作为系统中人的要素的总代表, 从微观层面到宏观层面考虑个体属性或团队属性; 作为用户代表, 与管理层和工程师建立良好的沟通协调界面。

(4) 学科目标协调。①要确保系统具有高效能, 包括(系统的)安全性、生产力、效率、有效性、质量、创新、灵活性、可靠性、可持续性等; ②要确保系统宜人, 即满足人的多层级需求, 包括安全与健康、满意度、审美、价值实现与个性发展等。因此, 人因工程应确保系统实现安全、高效和宜人的目标。当然, 确保所有目标要素同时

实现是不容易的,往往存在矛盾和挑战,在某些情况下需要进行取舍和权衡。例如,在速战速决的场合需要战斗员聚集力量发挥装备最大战力时,有可能需要降低一些舒适性要求。

文献[10]还指出人因工程的关键科学问题包括人的作业能力特性及其作用机制、人因失误/人因可靠性与安全性、人机交互新技术的基本原理、人因设计与测评方法、人机系统建模与仿真以及未来社会发展的人因方案等。

1.4 智能人机交互

从交互设计到人机交互再到人因工程,贯穿始终的一条主线就是如何使人有效地利用各种媒介与客观对象更好地交流或协作。交互设计是“艺术”,人机交互是“技术”,人因工程则是在更高层面、更大范围内的系统性统筹。无论如何设计人机界面,无论是人操作机器还是人通过机器与客观世界打交道,都有一系列共性的软硬件技术问题需要解决,其中一部分是通过计算机手段辅以智能算法使这种交流更加自然、舒适,这就是智能人机交互的研究范畴。

在智能人机交互中,“机”是一个泛化的概念,可以是台式机、笔记本电脑、平板电脑、虚拟现实设备、增强现实设备,也可以是手机、智能音箱、电视机,还可以是汽车、飞机、核电站、空间站,也包括连接不确定人群的各种云基础设施,只要不是人与人之间的面对面直接沟通,只要是通过计算机手段对这种沟通进行了智能化释义,从本质上讲都属于智能人机交互范畴。从信息流向上讲,智能人机交互包括人到机器的输出,以及机器到人的反向输出。

(1)人到机器的输出。传统手段主要是最早期的纸带,以及后来普遍使用的键盘、鼠标、写字板等,智能人机交互则主要包括语音/语义、手势/肢体动作、眼动、肌电、脑机接口等,能够使机器更加主动地去理解人的意图。语音/语义识别技术可以使机器理解人类的语言;手势/肢体动作识别技术可以使机器理解人类的动作;眼动识别技术可以追踪人类视线,用视线锁定目标;肌电采集与识别技术可以从肌肉群和神经信号中提取动作行为意图;脑机接口则尝试通过侵入手段或非侵入手段直接读取大脑的活动。除了人类主动向机器输出自己的意图信息外,机器也尝试在人类不自知的情况下解读人类,例如,MIT 媒体实验室的 Picard^[11]提出了情感计算(affective computing)的概念,通过分析人脸表情、身体姿态和语音语调等来识别特定人类个体的情绪情感状态。

(2)机器到人的反向输出。传统手段主要通过屏幕、声音向用户传达信息,智能人机交互则新增了虚拟现实、增强现实、生成式预训练模型、数字人、元宇宙等新型交互输出方式和解决方案。虚拟现实技术试图直接连接人类的视觉系统和听觉系统,将来自真实世界的视听通道信息源完全替换成计算机生成的虚拟世界信息

源,使人仿佛置身于一个全新的世界。为了更好地“欺骗”人类,虚拟现实技术还在尝试融合使用体感、触觉、嗅觉等新技术和用户交互,使用户获得更加逼真的沉浸式体验;增强现实技术致力于通过光学手段作用于人类视觉系统,尝试在真实物理世界之上叠加各种数字信息,使人观察到的是一个虚实叠加的世界;生成式预训练模型技术则在主动学习互联网海量信息的基础上具备了知识归纳推理和语言抽象输出能力,使机器像人类一样与人类进行自然交流;数字人技术将以假乱真的人类形象、行为习惯甚至个性差异赋予网络世界中的数字化智能体,使数字化智能体在虚拟世界披上人类的外衣,让人难分真假;元宇宙技术则尝试擘画未来整个万物互联世界的网络生态,通过构建一系列架空的数字世界使全人类以沉浸方式生活在其中,而这些数字世界必须以人类可直观理解的方式呈现,其中需要大量的智能人机交互技术介入。

智能人机交互技术的应用场景可以分为三大类,即生活娱乐类、生产设计类和行动任务类。

(1)生活娱乐类:主要应用于常用电子产品、高端电子娱乐、楼宇场馆展示等场合,根据需要的不同,主打人机交互的易用性、逼真度和震撼效果。

(2)生产设计类:主要应用于工业设计、艺术创作等场合,根据需要的不同,主打人机交互对复杂操控对象的全面支持性、准确度和表达度。生产设计类的工作具有创造性,设计软件的功能就必须具有足够的开放性。从交互界面上讲,这种开放性往往是以大量层叠的菜单选项或复杂的键盘鼠标输入呈现的,智能人机交互的任务是优化和简化这种交互方式。

(3)行动任务类:主要应用于复杂系统操作(如远程手术、遥控操作)、任务执行支持(如生产装配、设备维护)、军事行动等场合,根据需要的不同,主打人机交互的快速性、简洁性和容错性。智能人机交互在行动任务类的工作中往往以呈现工具和信息助手的方式出现,是服务于行动任务本身的,这就要求智能人机交互技术在其中不能喧宾夺主,向用户推送过多看似有用实则无用的信息是致命的,例如,很难想象医生在远程手术中,屏幕上突然出现一大堆术后用药信息提示。在这类场景中,智能人机交互的任务是在最需要的时刻用最快的速度推送最重要、最简洁的信息。当智能人机交互在行动任务类的工作中作为人类操作机器的输出工具时,同样要求人机交互操控的快速性、简洁性和容错性。

在国防领域,智能人机交互也变得越来越重要,在行动任务类的智能人机交互场景中,更加强调该人机交互是在特因环境下的人机交互。特因环境这一称谓在我国多出现于载人航天领域^[12],专指航天员在长期在轨飞行中,所处的微重力、昼夜节律紊乱、狭小空间、辐射等环境条件。在特因环境下,航天员往往会出现应激损伤,包括神经内分泌免疫功能紊乱、心血管系统紊乱、骨骼肌肉运动系统紊乱三大综合征。在军事行动领域,将特因环境这一概念进行了拓展,军事行动特因环境

专指作战人员在作战中所处的特殊环境,往往具备活动空间受限、感知条件受限、信息表达受限、行动能力受限、任务时间受限、强对抗烈度、心理高度紧张、失误不可修复等特点。在这种环境下,作战人员同样面临神经内分泌免疫功能紊乱、心血管系统紊乱、骨骼肌肉运动系统紊乱三大综合征的挑战。

在军事行动特因环境下,智能人机交互的基础软硬件支撑和交互范式都需要有专门的考量,主要体现在五个方面:多模融合、易于激活、按需推送、算通约束和自主容错。

(1)多模融合。受限于军事行动特因环境,日常生活中的语音识别、手势识别等技术不再适用。例如,对于语音识别技术,坦克驾驶员处于高噪声环境中,其指令很可能被噪声淹没,此时在语音识别技术的基础上引入基于图像的唇语识别技术就可以很好地解决这一问题。再如,对于手势识别技术,将惯性测量单元(inertial measurement unit, IMU)模块集成在手套上,将手套持续捕获的手部运动参数(角速度、加速度)与头戴式摄像头采集的连续视觉帧序列相融合,构建惯性传感数据与计算机视觉特征的多模态融合识别框架,可显著提升复杂场景下的手势识别鲁棒性。对于所有的战场人机交互环节,都应考虑作战人员在特因环境三大综合征情况下人机交互输出变形导致的交互失败,通过多模态融合识别的方式提高分类准确率。

(2)易于激活。处于军事行动特因环境下的作战人员由于注意力过度集中或精神紧张等因素会出现感知认知能力障碍。例如,紧盯敌方目标的作战人员会对周边其他环境的变化视而不见、充耳不闻;高度紧张的人面对即使最简单的提示信息都有可能出现精神短路、大脑一片空白的现象,当需要进行机器操作时,手忙脚乱,忘记如何唤醒系统。在设计智能人机交互时需要考虑上述因素,使操作尽可能简单、简化,并尽可能提供简洁的提示,甚至是机器自主的交互引导。

(3)按需推送。作战人员的目标是完成任务,智能人机交互起到辅助达成这一目标的作用,在使用智能人机交互技术的过程中必须始终牢记这一点。智能人机交互设计人员和工程师总是希望把能够提供的交互功能全部呈现给用户,让用户自己决定使用哪些交互功能,而过多的呈现和操作选择在战场上往往是致命的,因为人类大脑处理信息的能力有限,来自智能人机交互系统的冗余信息只会挤占为数不多的大脑处理能力,分散战斗人员的注意力。所以,在特因环境中使用智能人机交互技术时必须牢记“少就是好”的信条,只让最需要的交互功能在最需要的时刻以最简单的形式提供给使用者。

(4)算通约束。未来的智能化无人战场是一个电磁空间对抗的战场,通信能力一定会被极大地压制,云端无法保证提供稳定的计算和存储服务。对智能人机交互系统而言,必须要有离网计算能力,要有在通信、计算降级情况下的智能人机交互解决方案。