



人机情感交互

毛 峡 薛雨丽 著



科学出版社

人机情感交互

毛 峡 薛雨丽 著

科学出版社

北京

内 容 简 介

本书在详细介绍人机情感交互概念的基础上,分析了当前人机情感交互的研究前沿,特别是在情感模型、人脸表情交互、语音信号情感交互、肢体行为情感交互、生理信号情感识别、文本信息中的情感提取和情感仿生代理等方面的最新研究成果,使读者对人类与计算机之间的情感交互具有更深的理解,对促进我国在该领域的快速发展具有积极意义。

本书适合信息类、计算机类等专业研究生、高年级本科生阅读,还可供情感计算、人机交互和人工智能等相关领域的研究人员参考使用。

图书在版编目(CIP)数据

人机情感交互/毛峡,薛雨丽著. —北京:科学出版社,2011

ISBN 978-7-03-031799-5

I. 人… II. ①毛… ②薛… III. 人-机系统-研究 IV. TB18

中国版本图书馆 CIP 数据核字(2011)第 131875 号

责任编辑:张艳芬 / 责任校对:钟 洋

责任印制:赵 博 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2011 年 7 月第一版 开本:B5(720×1000)

2011 年 7 月第一次印刷 印张:15 1/4

印数:1—3 500 字数:292 000

定价: 55.00 元

(如有印装质量问题,我社负责调换)

本书获得以下基金支持

科技部“863”信息技术领域探索导向类课题“人机情感交互模式与技术研究”(No. 2006AA01Z135)

科技部中日国际科技合作项目“具有情感的智能仿生代理研究”(No. 2010DFA11990)

科技部中国与罗马尼亚政府间科技合作项目“复杂环境下的情感表情识别研究”(No. 39-5)

国家自然科学基金项目“基于情感仿生代理的多模交互标记语言研究”(No. 60873269)

国家自然科学基金项目“人机交互中的人脸情感表情识别研究”(No. 60572044)

教育部博士点基金项目“人机交互中的情感仿生代理研究”(No. 20070006057)

前　　言

人机交互是研究人、计算机以及人与计算机之间相互关系的技术。随着计算机技术的迅猛发展，人机交互技术也发生了质的变化。人机交互的发展过程是从人类适应计算机到计算机不断适应人类的发展过程，经历了语言命令交互、图形用户界面交互、人机自然交互到人机情感交互。近年来，人机情感交互已成为人机交互领域的研究热点，其内容涉及数学、心理学、计算机科学、人工智能、认知科学等众多学科。人机情感交互就是要赋予计算机类似于人的观察、理解和生成各种情感特征的能力，最终使计算机能够像人一样与人类进行自然、亲切、生动和富有情感的交互。

本书是一部关于人机情感交互基本原理和相关技术的著作，从人机情感交互的概念到当前人机情感交互的研究前沿，较为完整地介绍了人机情感交互所涉及的基本问题、主要研究成果和发展方向，将使读者对人类与计算机的情感交互有更深的理解。本书是情感计算领域一部较为全面的论著，深入剖析了该领域的研究动态，对促进我国在该领域的快速发展具有积极意义。

本书共 8 章。第 1 章为绪论，主要简述人机交互的概念、发展历史和趋势，以及人机情感交互的概念和研究内容。第 2 章为情感模型，主要介绍现有的若干种情感建模方法，分为基于基本情感论的情感模型、基于维度空间论的情感模型、基于认知机制的情感模型、基于个性化的情感模型以及其他情感模型，并举例介绍了情感模型的典型应用。第 3 章为人脸表情交互，主要介绍人脸表情建模、人脸表情识别、人脸表情识别的发展方向和人脸表情生成。第 4 章为语音信号情感交互，主要介绍大脑对声音信号的处理机制，语音情感识别以及情感语音合成。第 5 章为肢体行为情感交互，主要介绍手势情感识别、身体姿势情感识别以及身体姿势情感表达。第 6 章为生理信号情感识别，主要介绍情感的生理机制、生理信号情感识别及应用。第 7 章为文本信息中的情感，主要介绍基于词语的处理技术、基于数学统计的语言模型、基于语言理解的处理方法、基于语料库加工的处理方法以及文本情感提取方法。第 8 章为情感仿生代理，主要介绍移情仿生代理、基于仿生代理的标记语言以及情感机器人。

本书是北京航空航天大学电子信息工程学院模式识别与人工智能研究室全体师生共同努力的成果。自 1997 年以来，作者所领导的课题组长期从事情感信息处理及人机情感交互的研究，取得了大量研究成果，一并整理汇集到本书中，与读者共享。其中，毛峡教授负责章节设计、参与全书撰写和审校，第 1 章由毛峡教

授执笔,第2章由王晓侃执笔,第3章和第5章由薛雨丽执笔,第4章和第6章由陈立江执笔,第7章由易寒飞执笔,第8章由李峥执笔。特别感谢江琳、罗纳、王重平、吴星宇和魏鹏飞等研究生为本书的校正工作所付出的辛勤劳动。

本书撰写的内容得益于科技部“863”信息技术领域探索导向类课题“人机情感交互模式与技术研究”(No. 2006AA01Z135)、科技部中日国际科技合作项目“具有情感的智能仿生代理研究”(No. 2010DFA11990)、科技部中国与罗马尼亚政府间科技合作项目“复杂环境下的情感表情识别研究”(No. 39-5)、国家自然科学基金项目“基于情感仿生代理的多模交互标记语言研究”(No. 60873269)、国家自然科学基金项目“人机交互中的人脸情感表情识别研究”(No. 60572044)和教育部博士点基金项目“人机交互中的情感仿生代理研究”(No. 20070006057)等的研究成果,感谢有关专家对本书的推荐和鼓励,并向书中所有参考文献的作者表示感谢。

由于作者水平有限,书中欠妥和纰漏之处在所难免,恳请读者和同行不吝指正。

毛 峡

2011年3月10日

目 录

前言

第1章 绪论	1
1.1 语言命令交互阶段	1
1.2 图形用户界面阶段	2
1.2.1 “Memex”信息机器的构想	2
1.2.2 可直接构造图形图像的 Sketchpad 系统	3
1.2.3 世界第一只鼠标	3
1.2.4 使用图形用户界面的个人计算机	3
1.3 人机自然交互阶段	5
1.3.1 语音交互	5
1.3.2 普适计算	5
1.3.3 体感交互	5
1.3.4 基于视线追踪的人机交互	6
1.3.5 第六感交互	7
1.3.6 虚拟现实	7
1.4 人机情感交互阶段	9
1.4.1 人脸表情交互	9
1.4.2 语音情感交互	10
1.4.3 肢体行为情感交互	10
1.4.4 生理信号情感能识别	10
1.4.5 文本信息中的情感	11
1.4.6 情感仿生代理	11
1.4.7 多模情感人机交互	11
参考文献	12
第2章 情感模型	14
2.1 基于基本情感论的情感模型	14
2.2 基于维度空间论的情感模型	15
2.3 基于认知机制的情感模型	17
2.3.1 EM 模型	17
2.3.2 Roseman 情感模型	18

2.3.3 EMA 情感模型	18
2.3.4 Salt & Pepper 模型	19
2.4 基于个性化的情感模型	19
2.4.1 大五模型	19
2.4.2 Chittaro 行为模型	20
2.4.3 EFA 性格空间的构造方法	20
2.4.4 情绪—心情—性格模型	21
2.5 其他情感模型	22
2.5.1 Picard 的情感 HMM 模型	22
2.5.2 分布式情感模型	22
2.5.3 基于人工心理的状态空间情感模型	23
2.6 情感模型的应用	24
2.6.1 情感机器人	24
2.6.2 情感仿生代理	27
2.7 总结与展望	32
参考文献	32
第3章 人脸表情交互	34
3.1 人脸表情建模	34
3.2 人脸表情识别	39
3.2.1 人脸表情数据库	39
3.2.2 表情特征提取	46
3.2.3 表情分类方法	48
3.3 人脸表情识别的发展方向	54
3.3.1 鲁棒的表情识别	54
3.3.2 精细的表情识别	58
3.3.3 混合表情识别	60
3.3.4 非基本表情识别	61
3.4 人脸表情生成	64
3.4.1 人脸表情动画合成	64
3.4.2 智能人脸表情生成	69
3.4.3 眼动情感生成	77
3.5 总结与展望	84
参考文献	84
第4章 语音信号情感交互	93
4.1 声音信号的处理机制	93

4.1.1 大脑皮层下听觉中枢	93
4.1.2 大脑皮层听觉中枢	95
4.2 语音情感识别	96
4.2.1 应用前景	96
4.2.2 面临的主要困难	97
4.2.3 研究现状	99
4.2.4 情感语音数据库的建立	102
4.2.5 语音情感特征参数的提取	103
4.2.6 分类器的设计	112
4.2.7 情感分类器的融合	118
4.3 情感情音合成	122
4.3.1 语音合成及情感语音合成	122
4.3.2 基于韵律调节的情感语音合成	123
4.4 总结与展望	126
参考文献	126
第5章 肢体行为情感交互	129
5.1 手势情感识别	129
5.1.1 基于手套的手势识别	129
5.1.2 基于视觉的手势识别	130
5.2 身体姿势情感识别	132
5.2.1 身体姿势数据库	137
5.2.2 身体姿势特征提取	138
5.2.3 身体姿势识别方法	141
5.3 身体姿势情感表达	142
5.4 总结与展望	144
参考文献	144
第6章 生理信号情感识别	148
6.1 情感的生理机制	148
6.1.1 情感感知理论	148
6.1.2 大脑的情感中枢	150
6.1.3 与情感相关的内分泌腺	154
6.1.4 与情感相关的神经化学物质	155
6.2 生理信号情感识别	155
6.2.1 生理信号情感识别的研究意义	155
6.2.2 相关生理信号的基础知识	156

6.2.3 生理信号相关情感能识别的研究现状	161
6.2.4 生理信号情感能识别系统	163
6.3 应用系统	165
6.3.1 轻便式心电仪	165
6.3.2 情感监视手表	166
6.4 总结与展望	167
参考文献	167
第7章 文本信息中的情感	169
7.1 基于词语的处理技术	170
7.1.1 正向最大匹配分词	170
7.1.2 反向最大匹配分词	170
7.1.3 基于统计的词网格分词	171
7.2 基于数学统计的语言模型	171
7.2.1 隐马尔可夫模型	171
7.2.2 最大熵模型	172
7.3 基于语言理解的处理方法	173
7.4 基于语料库加工的处理方法	175
7.5 文本情感分析研究现状	178
7.5.1 词语的情感分析	178
7.5.2 句子的情感分析	179
7.5.3 篇章的情感分析	180
7.5.4 海量信息的整体倾向性预测	180
7.6 典型文本情感提取方法	181
7.6.1 基于关键词定位的文本情感提取方法	181
7.6.2 基于词汇关联的文本情感提取方法	181
7.6.3 基于统计的机器学习文本情感提取方法	182
7.6.4 基于规则的文本情感提取方法	182
7.6.5 基于常识的文本情感提取方法	187
7.7 总结与展望	188
参考文献	188
第8章 情感仿生代理	195
8.1 移情仿生代理	197
8.2 基于仿生代理的标记语言	203
8.3 情感机器人	225
8.4 总结与展望	228
参考文献	228

第1章 绪论

人机交互(human-computer interaction, HCI)是研究人、计算机以及他们之间相互关系的技术。人机交互界面是指人与计算机之间的通信媒体或手段,其模式已从语言命令阶段、图形用户界面交互阶段发展到自然和谐的感性用户界面阶段。以人为中心,自然、亲切和生动的交互成为发展新一代人机交互的主要目标。

人类之间的交流与沟通是自然而富有感情的,因此在人与计算机的交互中,人们也期望计算机具有情感能力。情感计算(affective computing)就是要赋予计算机类似于人的观察、理解和生成各种情感特征的能力,最终使计算机能够像人一样进行自然、亲切和生动的交互^[1]。

随着计算机技术的迅猛发展,人机交互技术也发生了质的变化,人机交互的发展过程也是从人适应计算机到计算机不断适应人的发展过程,它经历了以下几个阶段。

1.1 语言命令交互阶段

计算机发展的早期,人机之间的通信是通过机器语言完成的。计算机语言经历了从最初的穿孔纸带、汇编语言,直至高级语言的发展过程。这个过程也可以看做是早期人机交互的一个发展过程。

早期的人机交互是通过命令语言进行的,人与计算机之间通过语言的输入输出功能完成交互。最初,人机交互的方式是采用手工操作输入机器语言指令(二进制机器代码)控制计算机,这种形式很不符合人的习惯,既耗费时间,又容易出错,只有非常专业的人士才能做到运用自如。后来出现了FORTRAN、PASCAL、COBOL等计算机高级语言,使人们可以用比较习惯的符号形式描述计算过程,交互操作由经过一定训练的程序员即可完成。这一时期,程序员可采用批处理作业语言或交互命令语言的方式与计算机交流,虽然要记忆许多命令和熟练地敲击键盘,但已可采用较方便的手段来调试程序和了解计算机的执行情况。

20世纪60年代中期,命令行界面(command line interface, CLI)开始出现,通过这种人机界面,人们可以通过命令选择或命令语言方式进行人机交互。命令行界面可以看做是第一代人机界面。在这种界面中,人被看成操作员,计算机仅做出被动的反应,人只能通过用手操作键盘的方式输入数据和命令信息,界面输出

只能为静态字符,因此这种人机交互界面的自然性较差。命令行界面通常需要用户记忆操作的命令,因此不便于初学者使用。

1.2 图形用户界面阶段

图形用户界面(graphical user interface, GUI)的出现,使人机交互方式发生了巨大变化。图形用户界面是一种结合计算机科学、美学、心理学和行为学的人机系统工程。图形用户界面的主要特点包括桌面隐喻(desktop metaphor)、WIMP(windows icons mouse pointer)技术、直接操纵和“所见即所得(what you see is what you get, WYSIWYG)”。GUI简单易学,通过窗口、菜单、按键等方式就可以方便地操作,使不懂计算机的人也可以熟练地使用,从而拓宽了用户群,使计算机技术得到了广泛普及。

1.2.1 “Memex”信息机器的构想

GUI技术的起源可以追溯到20世纪40年代。1945年,美国科学家Bush(见图1.1)发表了一篇名为“As we may think”的论文,提出一种称为“Memex”信息机器的构想^[2],如图1.2所示。



图 1.1 Bush

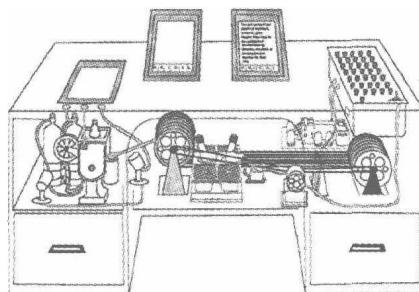


图 1.2 “Memex”信息机器的构想^[2]

Bush认为这种信息机器具有与书桌类似的外观,还有两个可触摸操作的显示器、一个输入键盘和一个扫描仪,通过这个信息机器,用户可以访问到人类的所有知识库。在那个年代,计算机远未产生,Bush的想法纯属科学幻想。但不可否认,该设想极富远见,它成功地描绘了半个世纪后的计算机形态,也许微软的平板计算机正是从Memex直接触摸操作概念中获取的灵感。

Bush将这种机器命名为Memex,也就是“memory extender”两个单词词首的组合,意思是“记忆的延伸”^[2]。

1.2.2 可直接构造图形图像的 Sketchpad 系统

20世纪60年代,美国麻省理工学院(MIT)的Sutherland(见图1.3)发明的Sketchpad系统可以在显示屏上直接构造图形图像,如图1.4所示。该系统首次引入了菜单、非重叠的瓦片式窗口和图标,同时可用光笔进行绘图操作,此发明代替了通过键盘向计算机输入代码公式的人机交互方式。更具革命性的是,在显示屏上作某些改动后,存储在计算机中的信息可以改变和更新。Sketchpad的成功奠定了Sutherland作为“计算机之父”的基础^[3]。



图1.3 Sutherland

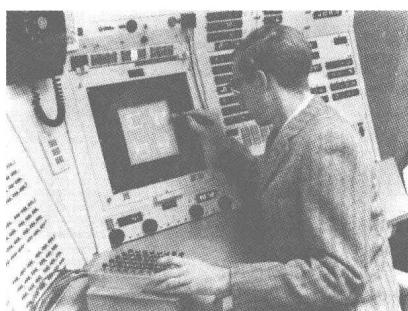


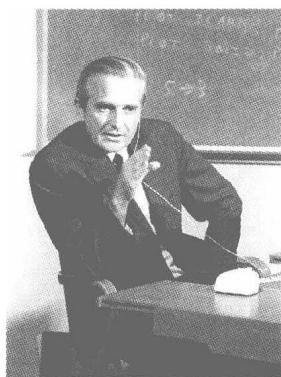
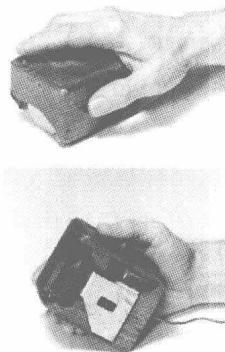
图1.4 Sutherland博士于1962年在演示最初的“画板”程序

1.2.3 世界第一只鼠标

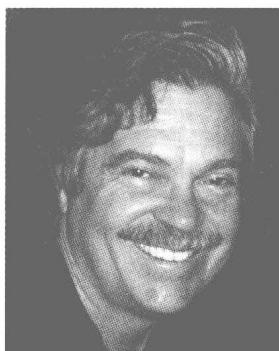
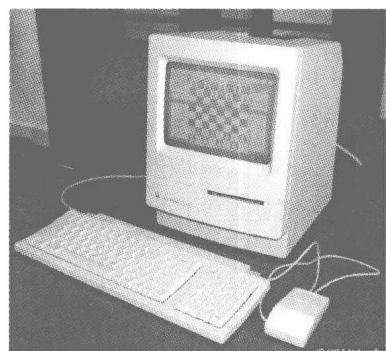
1968年12月9日,世界上第一只鼠标(mouse)诞生于美国斯坦福大学。它的发明者是年轻科学家Engelbart博士(见图1.5)。设计这只鼠标的目的是为了用鼠标代替键盘繁琐的输入,使计算机的操作更加简便。这只鼠标外形是一只手掌大的小木头盒子,如图1.6所示,其工作原理是由盒子底部的小球带动枢轴转动,继而带动变阻器改变阻值产生位移信号,位移信号经计算机处理后,屏幕上的光标就可以移动。从此以后,鼠标经过不断改进,在苹果、微软等公司的图形界面系统上得到了成功的应用,使鼠标与键盘共同成为目前计算机系统中必备的输入装置。特别是20世纪90年代以来,随着网络热在全球范围的升温,鼠标已经成为人们必备的人机交互工具^[4]。

1.2.4 使用图形用户界面的个人计算机

20世纪70年代,Kay(见图1.7)在施乐公司的PARC(Palo Alto Research Center)领导科研人员全力攻克图形化设计这一IT技术的战略制高点,并在1973年与同事构建了Alto个人计算机,如图1.8所示。Kay成为Macintosh和Windows的先驱。施乐公司在首台个人计算机Alto上首次开发了位图(bit mapping)

图 1.5 “鼠标之父”Engelbart^[4]图 1.6 世界第一只鼠标^[4]

显示技术,为开发可重叠窗口、弹出式菜单、菜单条等提供了可能。这些工作奠定了目前图形用户界面的基础,形成了以图形窗口(window)、图标(icon)、菜单(menu)和指针装置(pinting device)为基础的第二代人机界面,即 WIMP 界面。1984 年,苹果公司仿照 PARC 的技术开发出了新型的 Macintosh 个人计算机,如图 1.9 所示,将 WIMP 技术引入计算机领域,这种全部基于鼠标及下拉式菜单的操作方式和直观的图形界面引发了计算机人机界面的历史性变革^[5~7]。

图 1.7 Kay^[5]图 1.8 Alto 个人计算机^[6]图 1.9 Macintosh 个人计算机^[7]

与命令行界面相比,图形用户界面的人机交互自然性和效率都有较大的提高。图形用户界面很大程度上依赖于菜单选择和窗口小部件(widget)。经常使用的命令大都通过鼠标来实现。鼠标驱动的人机界面便于初学者使用,但重复性的菜单选择会给有经验的用户造成不便,他们有时倾向于使用命令键而不选择菜单。另外,图形用户界面需要占用较多的屏幕空间,并且难以表达和支持非空间性的抽象信息的交互^[8~10]。

1.3 人机自然交互阶段

人机自然交互可以理解为利用人类的日常交流方式与计算机进行交互。在人与人的交流中,人类可以利用语音、肢体、手势、眼神等方式实现交互。随着计算机技术、网络技术、模式识别技术以及虚拟现实技术的发展,采用上述方式与计算机交流并进行协同工作已经成为现实。

1.3.1 语音交互

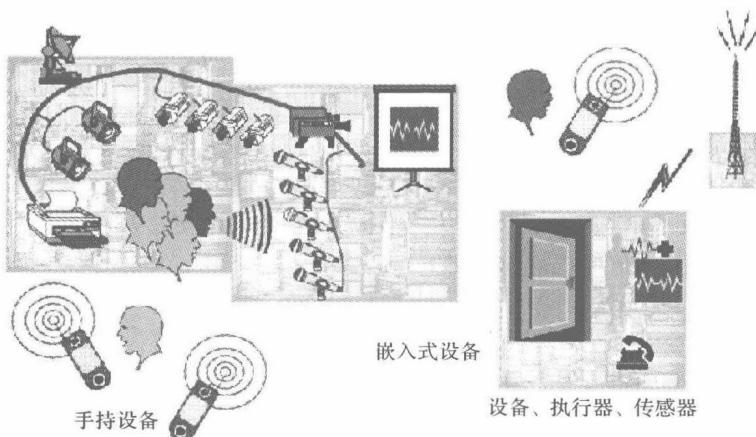
语音交互是人们日常生活中最常用的交互方式,用语音控制计算机并与其进行交互是人类一直以来所追求的目标。微软自1993年起就将语音技术作为自然人机交互的重要组成部分并专门成立了语音研究组。目前,随着微软新一代操作系统Windows Vista和Windows 7的推出,其新一代语音合成与语音识别技术,即Text-To-Speech和Speech Recognition技术也逐步进入千家万户。

1.3.2 普适计算

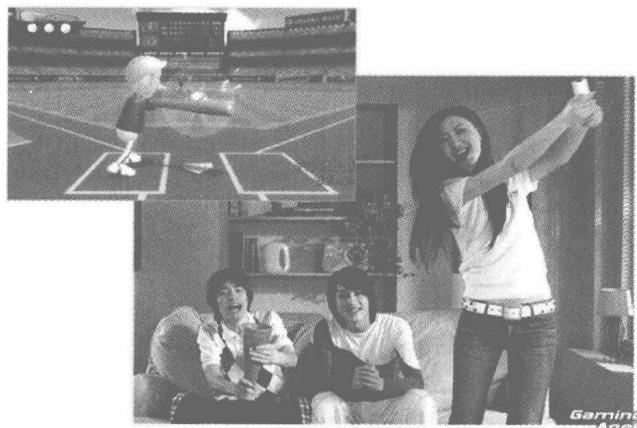
1991年,前施乐公司PARC首席科学家Weiser提出了普适计算(ubiquitous computing)的概念,这一概念强调和环境融为一体计算,计算机本身则从人们的视线里消失。在普适计算的模式下,人们能够在任何时间、任何地点、以任何方式与计算机进行交互。国际上已开展了许多普适计算的项目,其中较为著名的是美国麻省理工学院计算机科学和人工智能实验室所进行的氧气(oxygen)计划,其负责人Brooks解释oxygen计划为“关于以人为本的普适计算”,其目的是使丰富的计算和交互能力像空气一样无所不在、自由地融入人们的生活之中,该计划如图1.10所示。其他比较著名的项目还有美国斯坦福大学的交互工作空间(interactive workspace)项目,美国伊利诺伊大学的活动空间(active space)项目,以及微软的舒适生活(easy living)项目等^[11,12]。

1.3.3 体感交互

2006年,日本任天堂公司推出了家用游戏主机Wii,Wii与众不同的特色是它的标准控制器Wii Remote。Wii Remote的外形为棒状,如同电视遥控器,可以单手操作。除了像一般遥控器可以通过按钮进行控制,它还有两项功能:指向定位和动作感应。前者就如同光线枪或鼠标一样可以控制屏幕上的光标,后者可侦测三维空间中的移动及旋转,将两者相结合即可达成所谓的“体感操作”。Wii Remote在游戏软件当中可以化为球棒、指挥棒、鼓棒、钓鱼竿、方向盘、剑、枪、手术刀、钳子等工具,使用者可以通过挥动、甩动、砍劈、突刺、回旋、射击等各种方式来

图 1.10 麻省理工学院的 oxygen 计划^[11]

实现与游戏场景的交互,如图 1.11 所示。

图 1.11 使用 Wii Remote 的游戏场景^[13]

1.3.4 基于视线追踪的人机交互

艺术大师达·芬奇说过:眼睛是心灵的窗户。人们在观察外部世界时,眼睛总是与其他人体活动自然协调地工作,眼睛的视线信息可以反映人的心理状态,同时对心理状态的表达有着重要的辅助作用,因此眼睛的视线可作为人机交互的通道之一。

2008 年,日本东京大学与日本国立信息学研究所开发出基于视线追踪的人机交互系统(见图 1.12),计算机可实时感知用户的视线信息并做出智能反应。

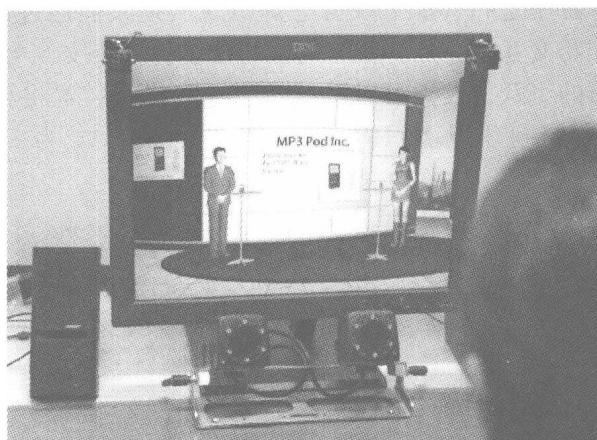


图 1.12 基于视线追踪的人机交互系统^[14]

1.3.5 第六感交互

在上百万年的进化中,我们一直在感知周围的世界。当遇到某事某人或来到某地时,我们使用五感来收集信息,然后决定采取什么样的行动。但是,大多数有用信息并不来自五感,而来自人类千百万年来对万事万物积累的各种数据、体验和知识,这些信息在互联网上已经逐渐汇聚成信息的海洋。虽然计算机设备的小型化技术已经可以让我们把计算机放进口袋,随时随地保持与数字世界的连接,但是仍然无法让数字设备直接参与我们与真实世界之间的互动。第六感技术跨越了这个鸿沟,将无形的数字信息和有形的世界连接了起来,让我们能够通过自然手势与这些信息交流。第六感系统通过无缝整合信息和现实世界,突破了对信息的限制,将整个世界放进了计算机中。

2009年,美国麻省理工学院发明了一种被称为第六感(sixth sense)的交互技术,应用该技术人们可以用双手操控一切与计算机进行实时交互。这套第六感系统由摄像头、投影仪、指环和计算机构成,如图1.13所示。摄像头可以捕捉手指的运动轨迹,识别交互命令;投影仪负责在需要的地方实时生成图像(如将手掌变成iPhone的屏幕电话,在手臂上投影手表)^[15]。

1.3.6 虚拟现实

虚拟现实(virtual reality, VR)是近年来出现的高新技术,也称灵境技术或人工环境,涉及计算机图形学、人机交互技术、传感技术、人工智能等领域。虚拟现实是利用计算机模拟产生一个三维空间的虚拟世界,生成逼真的视觉、听觉、触觉等感觉。让使用者如同身临其境,自然地观察三维空间中虚拟世界内的事物,作