



WILEY



Interactive Displays:  
Natural Human-Interface Technologies

# 实感交互

## 人工智能下的人机交互技术

[美] 阿钦蒂亚·K. 鲍米克 (Achintya K. Bhowmik) 主编  
温秀颖 董冀卿 胡冰 译



机械工业出版社  
CHINA MACHINE PRESS



# 实感交互：人工智能下 的人机交互技术

[美] 阿钦蒂亚·K. 鲍米克 (Achintya K. Bhowmik) 主编  
温秀颖 董冀卿 胡冰 译  
王亚楠 审校



机械工业出版社

过往的科幻现已成真,在人工智能时代我们与计算机、手机和娱乐设备的交互正在经历革命性的变化,基于触摸、手势、语音和视觉的自然人机交互正在逐渐替代使用键盘、鼠标和游戏手柄等的交互。显示设备也从单纯的显示设备转变为提供更具吸引力和沉浸式体验的双向交互设备。本书将深入讲解基于触摸、手势、语音和视觉等自然人机交互领域的技术、应用和未来趋势。

本书适合从事人机交互领域工作的研究、设计、开发人员,相关专业师生,以及人工智能时代下对人机交互未来发展趋势有浓厚兴趣的人士阅读。

Copyright© 2015 by John Wiley & Sons, Ltd

All Rights Reserved. This translation published under license. Authorized translation from the English language edition, entitled Interactive Displays: Natural Human - Interface Technologies, ISBN: 978 - 1 - 118 - 63137 - 9, by Achintya K. Bhowmik, Published by John Wiley & Sons, No part of this book may be reproduced in any form without the written permission of the original copyrights holder.

本书中文简体字版由 Wiley 授权机械工业出版社出版,未经出版者书面允许,本书的任何部分不得以任何方式复制或抄袭。

版权所有,翻印必究。

北京市版权局著作权合同登记 图字:01-2015-1416 号。

## 图书在版编目(CIP)数据

实感交互:人工智能下的人机交互技术/(美)阿钦蒂亚·K. 鲍米克(Achintya K. Bhowmik)主编;温秀颖,董冀卿,胡冰译.—北京:机械工业出版社,2018.3

书名原文:Interactive Displays: Natural Human - Interface Technologies  
ISBN 978-7-111-59782-7

I. ①实… II. ①阿…②温…③董…④胡… III. ①人-机系统-系统设计 IV. ①TP11

中国版本图书馆 CIP 数据核字(2018)第 087425 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策划编辑:林 楨 责任编辑:阎洪庆

责任校对:陈 越 责任印制:孙 炜

北京中兴印刷有限公司印刷

2018 年 6 月第 1 版第 1 次印刷

184mm × 240mm · 20.25 印张 · 472 千字

标准书号:ISBN 978-7-111-59782-7

定价:99.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线:010-88361066 机工官网:www.cmpbook.com

读者购书热线:010-68326294 机工官博:weibo.com/cmp1952

010-88379203 金书网:www.golden-book.com

封面防伪标均为盗版

教育服务网:www.cmpedu.com

## 原书序

人类对机器的交互式使用可以追溯到数千年以前。世界上第一台自动贩卖机可能是古希腊工程师 Philo 在公元前 220 年发明的。只需投入一枚硬币，它就会把称量好的肥皂传送到盥洗台上。这是一台带有擒纵机构的机械装置。这台颇为先进的机器无疑代表了当时最前沿的科技，但要说它对社会行为的发展产生了显著影响，这是值得商榷的。在往后的 2200 多年里，我们发现自己已经处于明显不同的境地。仅仅 50 年前，第一台电容触摸屏问世了；30 年后这项技术发展成熟，已经广泛应用在笔记本电脑、销售点终端这样的高端设备以及掌上消费市场；又一个 10 年过去了，如今的触控设备，至少是手持装置已经开始无处不在。

这有力地推动了本书的出版。本书由一组不同领域的高级技术专家联合撰写，涵盖了包括触摸、声音和视觉等主要交互形式。前两类互动形式将分两章分别讨论，视觉部分将在后五章阐释，主要关注已经问世或亟待问世的视觉科技特性。随后两章将介绍用来开发多模态交互显示的多种方法。本书结尾探讨如何呈现最为真实的 3D 立体图像。由于当前显示系统仅能保留高强度数据，而想要获得近似于人眼直接观察到的自然效果，就得通过保留已丢失的相位信息来实现了。

自此，读者会得出结论：本书全面回顾了当前新兴科技的发展。其实不尽然，因为我更想指出的是智能交互技术对社会带来的影响。虽然这些影响主要是积极的，但是也可能存在某些消极方面。这些都是公众关注的重要问题，因而值得辩论。积极的方面包括使用便捷；能够通过电脑或手机系统进行直观推理和预测；用简单的声音指令就能对复杂的结果进行讨论和管理；为身体不便的用户带来便利，并使其充分体验当前的各种产品，等等。然而消极的影响是，比起现在，通信系统会更广泛地侵入用户的生活。通话中的手机不仅早已被怀疑用来追踪定位，其未来系统还会更深入探测我们的行为模式。原则上，依靠计量生物数据来辨别身份的安全系统应当比目前的芯片和个人识别码技术更值得信赖，然而一旦前者的安全系统受损，可能泄露的安全数据将比后者要多得多。

未来人机交互体验的丰富程度，除非当前用户亲身使用，否则他们是难以想象的。过往的科幻现已成真。在技术创造的诸多可能性被供应商和用户采纳之前，有些问题必须讨论和解决。本书为此提供了多方面的素材和依据。对参与该主题的科学技术工作者和开发交互产品的参与者来说，这将是一本重要的书；对于有兴趣了解或需要了解交互技术会如何影响未来社会与人际行为的广大读者来说，本书也不容错过。

**Anthony Lowe**  
于英国 Braishfield

## 原书前言

什么是“人机交互显示”呢？我们将其定义为不仅能够在屏幕上显示可视信息，还能感知和理解人类行为、接收用户直接输入的显示器。能够“感觉”到手指触摸的触摸屏已经十分普遍，尤其是那些装配在移动设备和一体化计算机上的。现在，新增的类人传感与理解识别技术正在推进新型交互式显示器及系统的开发，使其能够在所处的3D空间中“看”“听”且“领会”我们的行为。

我们运用多感官和多模态界面模式来理解周围的客观物理世界，并与人们在日常生活中交流。这些都是通过无缝拼接包括触摸、声音、姿势、面部表情和凝视在内的多种交互模式实现的。我们如果想通过人机交互来获取社交交互的丰富内涵，就必须为这些设备上能够感知与领悟用户的输入与活动的技术。因此，增加多种自然用户界面能够使人类互动的体验更为真实。

我们与计算机交互的方式经历了最近几十年的变革，依靠鼠标和键盘作为输入工具的图形用户界面已取代传统的基于文本输入的命令式界面。而眼下，随着自然用户界面（通过触摸、姿势、语音等模式的人机交互）的兴起，我们正目睹着下一场技术革命的开始。实施人机界面模式的最终目标就是为用户呈现自然、直观、身临其境般的交互体验。虽然当前的技术局限使得设计师和工程师不得不有所妥协，致使部分目标仅能在完成某个特殊产品时实现，但是为了实现最终目标，我们在近几年来不断取得重大进展。

本书聚焦自然用户界面，对快速兴起的人机交互式显示领域内的技术、应用以及发展趋势进行了深度解读。第1章主要介绍人类感知和理解过程的基本要素，回顾了以触摸、声音和视觉感应推理为基础的自然界面技术，以及通过该技术实现的人机交互过程；随后各章深入每种输入与交互模态的细节，在实现多感官和多模态交互的目标过程中，对技术的基本原理及其在多种用户界面模式中的结合与应用展开细致的探讨；最后一章总结了基本要求和技发展现状，展望了未来有望实现的“真实的”3D交互界面及其带来的真实的、沉浸式的交互体验。

我向编辑 Anthony Lowe 致谢，是他发现了著书探讨交互式显示的必要性。我感谢对本书做出贡献的企业界和学术界专家，感谢 Wiley 出版社的员工对本书的支持。最后，谨以此书献给 Shida、Rohan 和 Ava，没有你们的鼓励和支持我无法开展并完成这个项目。

Achintya K. Bhowmik  
于美国加利福尼亚州

# 目 录

## 原书序

## 原书前言

## 第1章 交互式显示的感知、理解与自然

### 人机界面 ..... 1

- 1.1 引言 ..... 1
- 1.2 人类感知和理解 ..... 3
- 1.3 人机界面技术 ..... 7
  - 1.3.1 过往的输入装置 ..... 7
  - 1.3.2 触控式交互技术 ..... 9
  - 1.3.3 声控交互 ..... 10
  - 1.3.4 视控交互 ..... 12
  - 1.3.5 多模态交互 ..... 15
- 1.4 “真实”3D交互显示探索 ..... 17
- 1.5 结语 ..... 19
- 参考文献 ..... 19

## 第2章 触觉感知 ..... 22

- 2.1 引言 ..... 22
- 2.2 触控技术简介 ..... 23
  - 2.2.1 触摸屏 ..... 24
  - 2.2.2 按大小和应用对触控技术进行分类 ..... 25
  - 2.2.3 按材质和结构分类的触控技术 ..... 27
  - 2.2.4 按检测物理量分类的触控技术 ..... 27
  - 2.2.5 按感知能力分类的触控技术 ..... 28
  - 2.2.6 触控技术的未来 ..... 29
- 2.3 触控技术的历史 ..... 29
- 2.4 电容式触控技术 ..... 32
  - 2.4.1 投射电容式触控技术(编号1) ... 32
  - 2.4.2 表面电容式触控技术(编号2) ... 39
- 2.5 电阻式触控技术 ..... 43

- 2.5.1 模拟电阻式触控技术(编号3) ..... 43
- 2.5.2 数字多点电阻式触控技术(编号4) ..... 48
- 2.5.3 模拟多点电阻式触控技术(编号5) ..... 49
- 2.6 声波触控技术 ..... 51
  - 2.6.1 表面声波触控技术(编号6) ..... 51
  - 2.6.2 声学脉冲识别触控技术(编号7) ... 53
  - 2.6.3 色散信号技术触控技术(编号8) ..... 56
- 2.7 光学触控技术 ..... 57
  - 2.7.1 传统红外线触控技术(编号9) ... 57
  - 2.7.2 多点触控红外技术(编号10) ... 61
  - 2.7.3 摄像光学触控技术(编号11) ... 63
  - 2.7.4 玻璃光学触控技术(平面散射检测)(编号12) ..... 68
  - 2.7.5 视觉光学触控技术(编号13) ... 69
- 2.8 嵌入式触控技术 ..... 72
  - 2.8.1 外嵌互电容式(编号14) ..... 74
  - 2.8.2 混合互电容式(编号15) ..... 74
  - 2.8.3 内嵌互电容式(编号16) ..... 76
  - 2.8.4 内嵌式光感(编号17) ..... 77
- 2.9 其他触控技术 ..... 79
  - 2.9.1 压力感测(编号18) ..... 79
  - 2.9.2 组合触控技术 ..... 81
- 2.10 结语 ..... 82
- 2.11 附录 ..... 82
- 参考文献 ..... 83

## 第3章 用户界面中的声控式交互

- 技术 ..... 88

## VI 实感交互：人工智能下的人机交互技术

3.1 引言	88	3.10.6 知识呈现与推理	123
3.2 语音识别	91	3.10.7 监控	123
3.2.1 语言的本质	91	3.10.8 推荐阅读文献	124
3.2.2 声学模型和前端模式	92	3.11 问题解答	124
3.2.3 使语音对齐隐马尔科夫模型 (HMM)的过程	93	3.11.1 问题分析	125
3.2.4 语言模型	93	3.11.2 寻找相关信息	125
3.2.5 探索：以每秒1000个单词完成填字 游戏	95	3.11.3 解答与依据	126
3.2.6 训练声学 and 语言模型	96	3.11.4 呈现答案	126
3.2.7 为特定说话人识别系统调整发声和 语音模型	96	3.12 分布式语音交互架构	126
3.2.8 “标准”系统外的其他系统	97	3.12.1 分布式用户界面	127
3.2.9 性能	98	3.12.2 分布的语音及语言技术	128
3.3 语音识别的深度神经网络	98	3.13 结语	129
3.4 硬件优化	100	参考文献	130
3.4.1 低电量唤醒运算	101	<b>第4章 视觉传感与肢体动作交互 技术</b>	136
3.4.2 特定运算的硬件优化	101	4.1 引言	136
3.5 稳健语音识别的信号强化技术	102	4.2 图像技术：2D和3D	137
3.5.1 稳健语音识别	102	4.3 姿势交互	140
3.5.2 单通道噪声抑制	102	4.4 结语	146
3.5.3 多通道噪声抑制	104	参考文献	147
3.5.4 噪声消除	104	<b>第5章 实时3D传感与结构光技术</b>	149
3.5.5 回音消除	104	5.1 引言	149
3.5.6 波束形成	105	5.2 结构化图案汇编	150
3.6 声音生物计量	106	5.2.1 2D伪随机汇编	151
3.6.1 引言	106	5.2.2 二进制结构化汇编	152
3.6.2 声音生物计量面临的挑战	106	5.2.3 多进制汇编	153
3.6.3 声音生物计量的新研究领域	107	5.2.4 连续正弦相位汇编	154
3.7 语音合成	107	5.3 结构光系统校准	157
3.8 自然语言理解	110	5.4 数字条纹投射 (DFP) 技术下的3D传感 示例	160
3.8.1 混合主导对话	111	5.5 实时3D传感技术	162
3.8.2 预设和填值技术的局限	113	5.5.1 数字光处理 (DLP) 技术的 原理	162
3.9 多轮对话管理	116	5.5.2 实时3D数据采集	164
3.10 规划和推理	119	5.5.3 实时3D数据处理与可视化	165
3.10.1 技术挑战	119	5.5.4 实时3D传感实例	166
3.10.2 语义分析和语篇表达	120	5.6 人机交互应用的实时3D传感	166
3.10.3 语用学	121	5.6.1 实时3D面部表情捕捉及其人机交互 的意义	167
3.10.4 对话管理协作	122	5.6.2 实时3D身体部分姿势捕捉及其人机	
3.10.5 规划和再规划	122		

交互的意义 .....	167	7.10 技术发展最新水平 .....	206
5.6.3 人机交互意义的总结 .....	168	7.11 结语 .....	207
5.7 最新发展 .....	169	参考文献 .....	207
5.7.1 实时3D传感与自然2D彩色纹理 捕捉 .....	169	<b>第8章 凝视跟踪</b> .....	208
5.7.2 超高速3D传感 .....	171	8.1 引言和研究动机 .....	208
5.8 结语 .....	173	8.2 眼睛 .....	210
参考文献 .....	173	8.3 眼动仪 .....	212
<b>第6章 实时立体3D成像技术</b> .....	178	8.3.1 眼动仪的种类 .....	212
6.1 引言 .....	178	8.3.2 角膜反射法 .....	214
6.2 背景 .....	179	8.4 反对和障碍 .....	216
6.3 立体匹配算法的结构 .....	181	8.4.1 人为方面 .....	216
6.3.1 匹配成本计算 .....	182	8.4.2 室外应用 .....	217
6.3.2 匹配成本聚合 .....	183	8.4.3 校准 .....	217
6.4 特征分类 .....	184	8.4.4 精度 .....	217
6.4.1 深度估计密度 .....	184	8.4.5 点石成金 (Midas Touch) 问题 .....	218
6.4.2 优化策略 .....	185	8.5 凝视交互研究 .....	218
6.5 实施平台的分类 .....	186	8.6 凝视指向 .....	219
6.5.1 仅用CPU的方法 .....	187	8.6.1 解决点石成金问题 .....	219
6.5.2 GPU提速的方法 .....	187	8.6.2 精度问题的对策 .....	220
6.5.3 硬件执行 (FPGA, ASIC) .....	188	8.6.3 鼠标指向和凝视指向对比 .....	221
6.6 结语 .....	190	8.6.4 鼠标和凝视协调 .....	222
参考文献 .....	190	8.6.5 凝视指向反馈 .....	224
<b>第7章 飞行时间法3D成像技术</b> .....	194	8.7 凝视姿势 .....	224
7.1 引言 .....	194	8.7.1 凝视姿势的概念 .....	224
7.2 飞行时间法3D传感 .....	194	8.7.2 姿势检测算法 .....	225
7.3 脉冲飞行时间法 .....	196	8.7.3 执行凝视姿势的人类能力 .....	226
7.4 持续飞行时间法 .....	196	8.7.4 凝视姿势字母表 .....	226
7.5 计算方法 .....	197	8.7.5 姿势从自然眼动中分离 .....	227
7.6 精度 .....	199	8.7.6 凝视姿势的应用 .....	228
7.7 局限性与改进 .....	200	8.8 作为情境的凝视 .....	229
7.7.1 时差测距的挑战 .....	200	8.8.1 活动识别 .....	229
7.7.2 理论局限 .....	200	8.8.2 阅读检测 .....	231
7.7.3 距离混叠 .....	201	8.8.3 注意力检测 .....	232
7.7.4 多径与散射 .....	202	8.8.4 应用凝视情境 .....	233
7.7.5 功率分配与优化 .....	202	8.9 展望 .....	233
7.8 飞行时间法摄像组件 .....	203	参考文献 .....	234
7.9 标准值 .....	203	<b>第9章 感知用户界面的多模态输入</b> .....	237
7.9.1 光的功率范围 .....	203	9.1 引言 .....	237
7.9.2 背景光 .....	205	9.2 多模态交互类型 .....	237
		9.3 多模态界面 .....	238



## VII 实感交互：人工智能下的人机交互技术

9.3.1 触控输入	238
9.3.2 3D 姿势	245
9.3.3 眼动跟踪和凝视	249
9.3.4 面部表情	250
9.3.5 脑机接口	251
9.4 多模态集成策略	252
9.4.1 框架式集成	253
9.4.2 合并式集成	254
9.4.3 程序性集成	254
9.4.4 符号/统计集成	254
9.5 多模态交互的可用性问题	255
9.6 结语	256
参考文献	257

### 第 10 章 生物计量学中的多模态交互：

#### 技术与可用性挑战

10.1 引言	262
10.1.1 身份确认动机	262
10.1.2 生物计量学	263
10.1.3 多模态生物计量学的应用特征	263
10.1.4 2D 和 3D 人脸识别	264
10.1.5 多模态案例研究	266
10.1.6 适应于盲人对象	267
10.1.7 本章结构	268
10.2 对移动生物计量平台的应用剖析	268
10.2.1 面部分析	268

10.2.2 语音分析	271
10.2.3 模型适应	272
10.2.4 数据融合	273
10.2.5 移动平台实施	274
10.2.6 MoBio 数据库和协议	275
10.3 案例研究：为视觉缺陷者进行可用性研究	276
10.3.1 头部姿势变化对性能的影响	276
10.3.2 用户交互模块：头部姿势质量评估	278
10.3.3 用户-交互模块：音频反馈机制	280
10.3.4 视觉缺陷者的可用性测试	282
10.4 讨论与结语	284
参考文献	285

### 第 11 章 迈向“真实的”3D 交互

#### 显示器

11.1 引言	287
11.2 生物视觉的起源	289
11.3 光场成像	294
11.4 迈向“真实的”3D 视觉显示	300
11.5 与 3D 显示屏上的视觉内容交互	308
11.6 结语	310
参考文献	311

附录 缩略语	313
--------	-----

# 第1章

## 交互式显示的感知、理解与自然人机界面

Achintya K. Bhowmik  
美国英特尔集团

### 1.1 引言

如今，可视化显示设备已成为丰富多彩的电子产品中不可或缺的一部分。作为人与电脑、通信系统和娱乐系统交互的主要界面，其应用已经融入居家、工作或出行等生活的方方面面。无论是腕上的手表，还是随身装在口袋或钱包里的手机，抑或是用来网上冲浪、获取多媒体信息的平板电脑，再或者是工作的笔记本电脑或台式电脑，还有客厅中心的巨屏电视、商务会议使用的演示投影仪，可视化显示器都是这些设备面向我们用户的“颜面”。

这类显示器频繁应用于各种特定的公共场所，比如机场自助登机手续办理终端，零售店自助付款机、大型购物商场的广告牌以及博物馆的公共展示——用途不计其数。近十年来，巨大的应用潜力和市场需求促进了全球可视化显示技术的研发。从移动显示到巨屏显示，多样化的产品层出不穷<sup>[1-5]</sup>。

只要扫一眼可视化显示设备的市场规模，我们就能很快领会它给生活带来的影响。来自显示产业分析公司 IHS 的报告说明，近五年来，销往世界各地的平面显示设备总额高达 170 亿美元<sup>[6]</sup>，年度出货量超过 50% 的增速也说明了这一技术的快速普及率。

总体来说，一台电子设备主要完成三项基本功能：接受用户指示，按照指示及所获信息执行某些处理功能，呈现输出或向用户报告处理结果。比如，当作者在笔记本电脑上进行本章的写作时，他首先用键盘和鼠标输入信息，然后微处理器就会执行文字处理软件，将敲击键盘和点击鼠标发出的命令转换成目标文本和格式，最后，电脑的液晶显示屏就会以可视化的输出实时显示文字。由此可见，设备里的显示子系统已经在向用户呈现信息方面发挥了至关重要的作用。除了某些特例之外，大多数近期生产的电子产品都配备了显示屏幕，唯一的目的是为了显示视觉信息。

然而近几年来，人机互动和用户界面范式一直在经历着快速的演变和创新。我们与电脑

## 2 实感交互：人工智能下的人机交互技术

交流的方式经过几十年的变革已经大不相同。在文本型的老式命令输入界面被淘汰以后，取而代之的是依靠鼠标和键盘输入的图形用户界面。随着更多自然用户界面的出现，下一场变革的帷幕正在我们的眼前拉开。未来，人机交流不但可以通过触摸、肢体动作、声音、表情和视线来实现，甚至还可以通过我们的思想！

我们正在不断研发高级传感器、系统、运算规则以及应用程序，以实现更为生动自然的互动体验。在这个过程中，运算装置除了能够把握交流意图之外，还能理解用户的表达与情感。这些兴起的界面技术和接踵而至的新型应用产品为显示技术乃至整个电子消费产业创造了振奋人心的机遇。随着自然用户界面的不断整合，显示设备也从以往视觉内容的单向显示转变成了可以接收用户输入的双向互动，这就推动了交互应用程序的开发和沉浸式体验的实现。触摸屏和触控优化界面以及各类应用产品的激增又把这场变革蔓延到了移动显示设备，自然界面技术由于其交互性的强化而不断延展，必然会重新定义整个显示技术和显示系统的维度。

本书全面解析了促使高度交互显示与显示系统兴起的自然人机界面技术与应用。那么什么是“人机交互式显示”呢？我们将其定义为不仅可以在屏幕上显示可视信息，还可以感知和理解人类行为并接收用户的直接输入。一旦装配上类似自然人的感知和理解技术，一个“真实”的人机交互式显示器就能“感受”并探测到我们的触摸，“听到”并回应我们的声音，“看到”并辨识出我们的面貌和表情，“理解”并阐释通过移动手指或其他身体部位发出的肢体指令，甚至能够根据语境推理出我们的意图。

虽然这些目标看起来非常远大，但是正如图 1.1 所示，依靠简单直观的自然人机界面，多种形态因素和应用系统加之自然用户交互技术已经对市场带来了巨大的影响。本书的讨论也在不断揭示这种影响，我们在自然感知、推理技术、系统整合和应用发展方面取得的重大进步将为人机交互的全面创新打下坚实的基础。

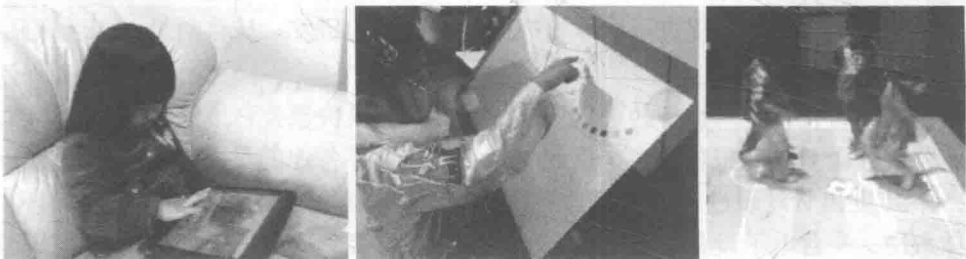


图 1.1 各种形态的交互显示器与应用系统已经占据了大片市场，如前面例子所述。除了传统意义上对用户显示视觉信息之外，许多系统内的显示器在直接人机界面设备中发挥着新的作用

图 1.2 描述了交互显示系统的通用功能模块及其流程。用户和显示系统的互动是受各个界面发出的指令支配的，也就是在开始和结束部分显示的输入和输出模块。输入模块由一组

传感器组成，能够把用户输入的物理刺激转换成电子信号。而输出模块则以物理刺激的形式，让用户感知并理解系统反向回应用户的行为。中间的模块处理必要的信号并执行运算功能以促进交流。

本章首先综述了人类感知和理解的基本原则，特别关注了我们在与物理世界的日常互动中部署的机制和流程。以此为基础，我们随后概述了运用自然界面技术（包括触摸、声音、视觉感知和互动）的人机互动过程，并简要梳理了史上最为成功的界面技术。接下来，我们将深入到每类输入与互动的模态细节，对技术原理及其在自然人机界面模式的应用，以及综合互动技术在实现直观的多感官、多模态互动方面的作用进行深入的探讨。本书最后一章总结了基本要求和科技发展现状，展望了未来有望实现的“真实”的3D交互式显示及其带来的真实的、沉浸式的互动体验。

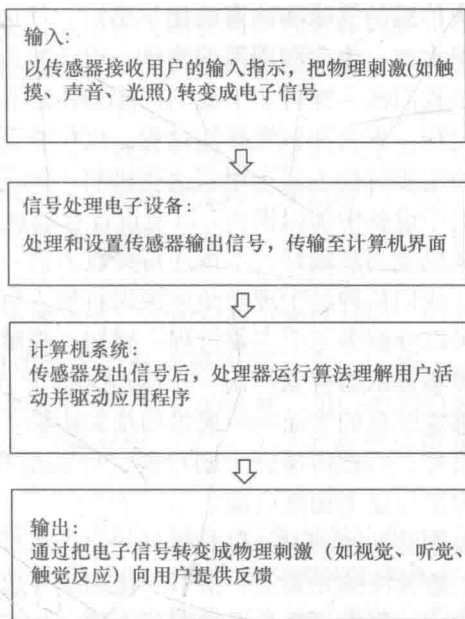


图 1.2 交互显示系统的功能框图。输入模块与输出模块指挥用户与显示器的互动，而信号处理与运算法则促进了这些互动的开展

## 1.2 人类感知和理解

本书伊始就提出实施人机交互界面方案的最终目标是让用户获得自然、本真和沉浸式的互动体验。虽然目前技术的局限性让设计师和工程师不得不做出妥协，仅能实现某些特定产品的部分目标，但是我们一直在实现总体目标的方向上取得进步。

这里需要进一步阐明一下。所谓“自然”，意思在于运用我们的自然机能与机器实现交流和互动。我们运用多感官、多模态的界面方案来理解周围环境和相互交流，将包括声音、表情、凝视、手势和肢体语言、触觉、嗅觉和味觉等在内的多模态互动无缝衔接。如此，创建自然界面就能使真实的生活体验融入人机互动之中。

所谓“本真”，意指该界面依靠我们多年养成的社交习惯而设计，仅要求用户使用最少的（理想是不需要任何）学习成本就能与机器进行交流。

所谓“沉浸式”，是一种真实世界与虚拟世界边界模糊化的体验，其中电脑或机器成为我们身体与大脑的延续，帮助我们完成任务。这是个很高的要求，需要几十年的持续研发才能接近这些目标。我们努力了解生动逼真的人机界面和交互方案，就能使我们以史为镜，了解人类——毕竟我们是“人机互动”这个词组的第一个字！

我们人类已经进化成了高等交际物种，受助于一个精干的大脑和一系列复杂的感知器官，包括丰富的视觉感知系统、听觉能力、接触敏感的皮肤和触觉感知，还要算上经过鼻腔

## 4 实感交互：人工智能下的人机交互技术

和舌头传感的气味和味道的化学感知。超过一半的人类大脑致力于处理感知信号，让我们能够认识太空、生命和周围的物体，也让我们在自然、本真的感知情境中彼此互动。

让我们深入探讨一下我们的感知传感和推理过程，即眼睛和视觉感知过程，耳朵和听觉感知过程，皮肤和触觉感知过程。仅仅专注于这三种感知模态的一个原因是我们与物理世界交互的实质过程主要运用到这些机制，而且我们也将看到，这些机制的功能能够依靠高新技术在电子设备中加以模仿，以便设计和制造高级互动显示器和系统。在人机交互中实现嗅觉和味觉机能当然最好，不过还得等技术进一步发展。

让我们从神经生理学角度探讨自然人机界面与交互显示系统，如图 1.1 所示。这个交互过程可以分解为三个主要过程：感知，理解和辨识，以及行为。从人的视角看，感知过程包括：搜集显示器视觉产出——通过光波介入人眼；说话人听觉产出——以声波形式介入人耳；感觉屏幕的表面——通过用指尖碰触。这些感知传感器将物理刺激通过传导过程转换成神经信号，后被传递到大脑皮层，也就是我们能够理解到“看”“听”和“触”的发生，随后辨识与思考相继启动。

根据感知和辨识过程的结果，我们将指令我们的身体行为。比如，我们把视线聚焦到显示器上想关注的元素上，指引手指触摸并启动屏幕上的具体内容，调整我们对声音产出的听觉注意力，摆出一个合适的面部表情，甚至用我们的手指和手来做一个动作。

我们首先综述一下视觉感知过程。我们仅关注与随后讨论密切相关的操作交互显示器的内容，并把其他更为详细介绍人类感知<sup>[7,8]</sup>的读物介绍给有兴趣的读者。人眼是人类进化的奇迹，特别体现在其构造上的极端复杂性，功能的有效性及其在连接感知世界与大脑枕叶视觉皮层方面所发挥的核心作用。如图 1.3 所示，人眼和相机的某些核心结构十分相似，都是

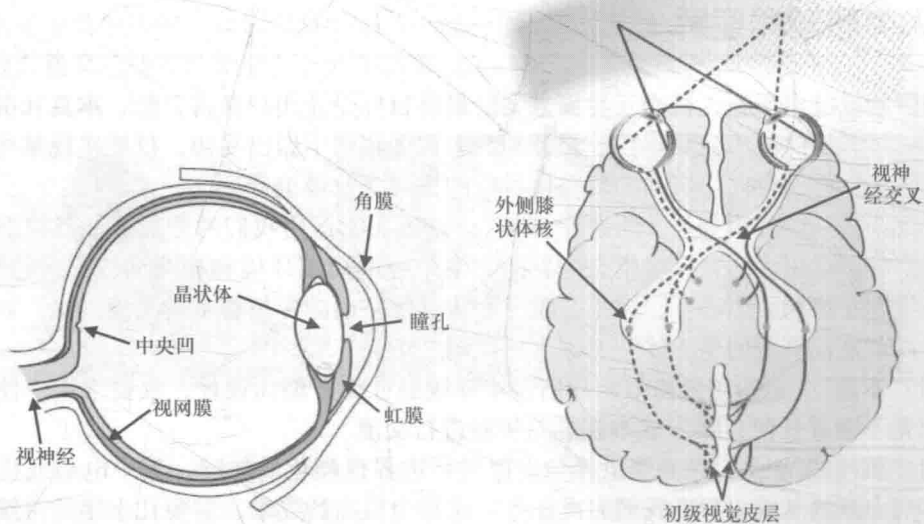


图 1.3 左：人眼解剖图。右：人的视觉系统使用双目成像方式。左视野是由双目的右侧感知到的，并映射到视觉皮层内的主要接收区域的右半部，右视野则经由另一条路线。观测物与双眼的距离是由双目视差察觉的，其他的视觉线索还包括动态视差、视差映射和焦距等

通过透镜系统把外景光源聚焦在眼部后方的视网膜上成像；视网膜周围含有称为感光器的感光细胞。眼部有两种类型的感光体，即有色觉的视锥细胞和无色觉的视杆细胞，后者能把光转换为神经信号。

这台相机的分辨率以及和处理器通信的带宽如何呢？视网膜包含大量的感光器——大约每只眼睛有 800 万个视锥细胞和 12000 万个视杆细胞——然而视觉体系却能够巧妙地发出景物在空间和时间上变化的信号，而不是由感光器探测到的绝对光强，以保持眼睛和大脑的通信带宽降到实际水平上。

当我们把目光投向一个物体且图像形成于视轴周围的一个相对较小的区域时，中心视觉的视敏度是最高的。这是因为视锥感光器最集中地分布于视网膜内的一个小区域——中央凹，这些感光器映射到视觉皮层内的一个比视网膜其他部分要大的区域。另一个相机的重要特质是光敏的动态范围，人眼的视觉跨径可达 10 个数量级，远远超过了现代数码相机的能力。

每只眼睛都是一部优秀的相机，像这样的相机我们拥有两部。人类的视觉系统包括 3D 和深度理解能力，有着双目成像方式以及其他诸如动态视差、视差映射和焦距等视觉线索，这些能让我们在 3D 空间内十分轻松地找到方向并于各种物像交互。双目成像已经普遍演化成大多数生物系统的特征。近期的化石研究论证其早在 5 亿多年前节肢动物生活的早寒武纪时代就已经存在<sup>[9]</sup>。强大的视觉系统的出现被认为是引发寒武纪大爆炸变革的导火线<sup>[10]</sup>。部分重叠的横向位移视野导致了“双目视差”，也就是由单眼捕捉到物体相对于另一只眼睛发生了横向位移。我们随后将会了解到，双目视差与观测物到观察人的距离成反比。

有这样的视觉系统帮助理解距离，猎物就更容易发现逼近的猎人而逃生，猎人也有更好的时机三角测距猎物的位置并实施捕猎。双目视觉因此被推定为生物进化成功的推动力，也是最早的哺乳动物的特质之一。时至现代，我们运用我们复杂的双目视觉系统来与 3D 世界互动。图 1.3 也简化地展示了将眼睛连接到视觉皮质的感觉传导路径。

接着，我们来思考一下听力感知的重要元素，包括耳朵和各个听辨过程。恰如眼睛，人的耳朵也有着精致的构造以及像声音传感器这样令人惊叹的功能。我们天然的麦克风——耳朵——能够感知超过 12 个数量级的声音强度以及 3 个数量级的音频（20 ~ 20000Hz）！

如图 1.4 所示，耳廓决定了气流携带声音信号进入含有耳鼓膜的耳道的方向。压力振荡经由中耳组织——锤耳、砧骨和镫骨得以放大，这些部位是人体拥有的最小骨头，英文中可分别用意为锤子（hammer）、铁砧（anvil）和马镫（stirrup）的单词表示，暗指它们是如何放大并向内耳部分传递声音信号的。最后，振荡声波被转经由神经冲动转换成神经信号，更具体地说是由位于呈收敛螺旋状的耳廓部位的听毛细胞转换的。这些神经信号随后发射到位于颞叶的大脑听觉皮层并被处理成能够感知的信号。

正如人眼一样，我们还有一对能在频率信号之外启动双声道感知方案的天然“麦克风”，它可以在 3D 空间内准确定位声音的来源。双耳 3D 感知以及极高的声压灵敏度对我们的进化过程十分重要，在日常生活中，它也对帮助我们在 3D 物理世界的穿梭和交流起到了不可或缺的作用。图 1.4 简单展现了位于人耳与大脑听觉皮层之间的神经分布路径。

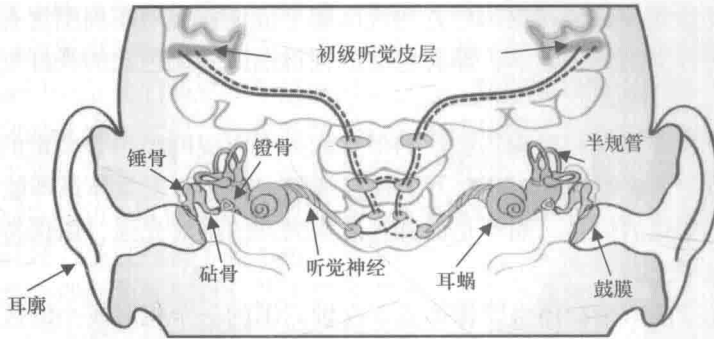


图 1.4 人耳和双声道构造的解剖图，简化地描绘了位于内耳耳蜗和大脑颞叶内的听觉皮层之间的神经分布路径。双耳信号以及频率信号被用来定位声音信号的来源

最后，我们再看看触敏性和触觉感知过程。触觉的感知过程又称皮肤感知，开始于皮肤内的机械性感受器，它们能够在相应的皮肤区域感受到因接触而产生的机械压力。图 1.5 描绘了 4 种主要的机械性感受器。视觉（眼睛）和听觉（耳朵）感知器官位于颅骨内，具有离大脑皮层相对较短的神经生理路径，而触觉感知器官（皮肤）却覆盖了整个身体。因此，来自触觉接收器的信号常常需要经过较长的距离（比如从手指到头部）。脊髓对触觉感受器来说就起到了“信息高速公路”的作用，把从接收器获得的信号传递到顶叶内的大脑体觉皮层——这部分大脑位于处理触觉过程的头部顶端区域。

神经外科医生 Wilder Penfield 在 20 世纪 50 年代关于触觉敏感的重大发现已经证明了人体邻近部位对大脑皮层邻近区域的映射<sup>[11]</sup>。更有意思的是，这项映射研究确立了作用于身体各个部分的大脑体觉皮层的相对比例。图 1.5 所示的“皮层矮人”（cortical homunculus）

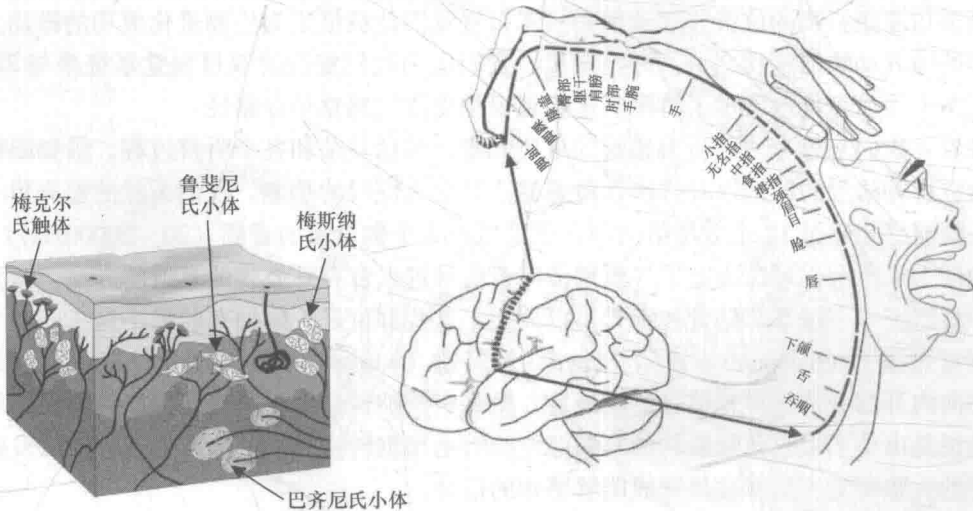


图 1.5 左：人体皮肤的解剖图。所示的主要四种类型的机械性感受器将触碰刺激产生的机械压力转变成神经信号。右：矮人皮层理论，由 Wilder Penfield 首次提出，后续学者陆续完善，揭示了体觉皮层处理来自身体各个部分的触觉信号的位置和相对比例。来源：[http://www.intropsych.com/ch02\\_human\\_nervous\\_system/homunculus.html](http://www.intropsych.com/ch02_human_nervous_system/homunculus.html)，转载获得 R. Dewey 的许可

的概念就是该理论的集中体现。别错以为这只是幅随意的讽刺漫画，其实这个矮人图呈现了一个人体各部分所占大脑体觉皮层的相对空间的比例模型。如图所示，该皮层组织致力于处理来自手指碰触的信号远超过其处理来自整个手臂和手腕，这恰恰证明了触摸屏用户界面设计师期望大量运用手指来实现触控式人机交互的合理性！

正如前面讨论过的，神经生理学对我们的感知过程有一个普遍的解释。感知系统的设计十分巧妙，绝大部分的大脑皮层组织与感知接收器最重要的部分是相连的。比如，视网膜中央凹与中心视力，耳蜗听毛细胞与听觉，手指尖与触觉等。虽然我们也拥有其他感觉机制，但是在与周围物理世界的交互中，我们更主要依靠的是看、听和碰触。因此本书主要关注眼睛、耳朵和触感作为自然人与显示器设备交流的主要模式。

相比起生物系统，当今大多数的计算和娱乐设备具有非常初级的感知和处理能力。就手机、平板电脑和笔记本电脑来看，它们是典型的“单眼”工作（仅有一个相机），就像希腊神话中的独眼巨人库克罗普斯一样。此外，它们大多数是单耳结构（仅有一个麦克风），还有许多尚未实现触敏（触摸屏），尤其是笔记本电脑。

但随着技术在多方面的迅猛发展，这一情况将有望在不远的未来得到改善。向自然和人类世界学习，工程师和设计师现在已经开始对计算和通信设备加入“类人”的感触和感知属性，让它们能够“看”“听”和“理解”人类行为和指示，并发挥这些功能以促进自然的、本真的互动。这些发展保证了人机交互实现超越键盘、鼠标、操纵杆和远程遥控的突破，并允许基于碰触、视觉与言语感知和识别技术的自然交互的使用。

尽管现实中我们每时每刻感知和洞察周围世界是那样的自然和随意，但是只有我们尝试在机器中实施这些感知功能的时候才能理解这些任务的复杂性。在下一节，我们将综述人机界面与电子设备的重要技术，包括最近几十年广泛采用的技术先例以及新近实现的与显示器和系统交互的自然本真模式。

## 1.3 人机界面技术

### 1.3.1 过往的输入装置

在深入讨论最新自然界面技术和由其推动的应用与用户体验之前，很有必要回顾一下历史，思考一下最成功的用户输出技术的发展。最近几十年间，随着我们在生活中的实践和接触，该技术已经成为人机交互技术的核心支柱。我们无意对人机界面技术及其相关的历史发展进行完整记述，想要在有限的章节内完成这项任务也不可能。我们现在能做的是把过去出版的许多综述文献介绍给有兴趣的读者<sup>[13-15]</sup>。

以下我们简述几项已被大众采纳的创新发明和主流产品，它们定义了时至当代的人机交互的主要方式。回首过往，我们感谢它们取得成功的重要因素——简易的技术应用，以最适度的价格使用最优的现有技术元素，尤其是应用某项发明来满足用户对丰富个人生活和各项活动的需求。



首先，无处不在的遥控装置可以说定义了我们与电视荧屏的交互关系并塑造了我们的内容浏览行为。尽管远程遥控的概念早在 1898 年就由 Nikola Tesla<sup>[16]</sup> 提出，第一个电视机遥控器却是由 Zenith 无线电公司于 1950 年开发并投入市场的，并形象地将遥控器命名为“懒骨头”<sup>[17]</sup>。

电视机自 20 世纪 20 年代起就已经风靡市场，但那时人们需要走到它跟前来调整控制旋钮，尽管自然的观看姿势就是坐在它面前的沙发上。因此，当时的环境对发明远程遥控来说是成熟的；需求很明确，而且技术水准也已经到位。Zenith 的“懒骨头”手持遥控器与电视机之间是有长线相连的。虽然这解决了人们的合理需求，使人们不离开沙发就能换频道，但避免不小心被电线绊倒的需求还是指向了无线遥控的发明。

到了 1955 年，也是由 Zenith 生产的“闪光助手”（Flash-matic）问世。通过光束指向分布在电视机屏幕四角的感应器，用户就能使用这款遥控器实现无线控制。这种激动的心情可以从宣传当日的杂志广告中体会：“不得不让你眼见为实！”虽然兴奋难抑，但是这款光控设备并不能在光亮的房间内很好地发挥作用，因为外界的光线会偶尔改变设置。Zenith 把下一代的设备更换成了超声波作为远程通信媒介，并命名其为“太空司令”，这才解决了问题。一则 1957 年的广告（见图 1.6）振振有词地宣扬了这款“坐享舒适沙发，无声遥控电视”的神奇体验。



图 1.6 1957 年的一则 Zenith “太空司令” 遥控装置的广告。来源：[www.tvhistory.tv](http://www.tvhistory.tv)（已取得引用许可）

此后的现代遥控技术发展更是日新月异，产品不仅融入了各种时尚精巧的形状元素，还安装了红外光以遥控娱乐装置。近几年还不断新添了动作感应和声音控制技术特征。

接下来，我们再回到 Douglas Engelbart 发明电脑鼠标的 1963 年，这是标志着人机交互新纪元的开始之年。在发明和装配鼠标之前，早期电脑输入局限于基于文本的键盘敲击指令。图 1.7 展现了第一个由 Engelbart 和 Bill English 构建的鼠标原型。鼠标由两个在互为直角方向上滚动的滑轮组成，随着鼠标在平面拖动，两个滑轮能跟踪鼠标在 2D 平面上的位置<sup>[18]</sup>。Engelbart 在 1961 年设想到该装置时正在参加一个电脑作图会议，思考着如何能构建一个能与电脑绘图对象简易高效互动的系统<sup>[19]</sup>。

值得注意的是，鼠标仅仅是 Engelbart 和他的斯坦福研究学院团队发明的众多电脑输入