



# 多媒体接口设计

Meera M. Blattner  
Roger B. Dannenberg 编

张知奋 龙利工 等译  
胡传国 审校

上海科学普及出版社

# 多媒体接口设计

Meera M.Blattner  
Roger B.Dannenberg 编

张知奋 龙利工 等译

胡传国 审校



上海科学普及出版社

9510060

## 内 容 提 要

本书介绍多媒体和多模态用户接口设计技术；叙述接口的模型，构成和设计，理论基础和实现技术，以及体系结构；涉及多媒体接口设计中许多重要问题，如：声音的各种表达形式，多模态输入和手势的应用，随时间变化的各种媒体（视频、动画、声音、手势和体位等）及其输入设备的设计，多媒体接口的智能性问题，多媒体图符化接口，对来自各种输入设备的二义性输入进行选择 and 校正的技术，体系结构的原理和理论，多媒体操作系统和对话管理，多媒体应用实例等。本书内容富有独创性和启发性。

读者对象：计算机专业技术人员，多媒体技术研究和开发人员，多媒体系统用户，大专院校有关专业师生。

(沪) 新登字第 305 号

责任编辑 田 文 胡名正

### 多媒体接口设计

Meera M. Blattner  
Roger B. Dannenberg 编

张知奋 龙利工 等译

胡传国 审校

上海科学普及出版社出版  
(上海曹杨路 500 号 邮政编码 200063)

---

新华书店上海发行所发行 上海市印刷七厂一分厂印刷

开本 787×1092 1/16 印张 22.25 字数 579000

1994 年 10 月第 1 版 1994 年 10 月第 1 次印刷

---

ISBN 7-5427-0800-7 / TP·189 定价: 平 装 32.00 元  
精 装 35.00 元

# Multimedia Interface Design

Edited by

Meera M. Blattner

Roger B. Dannenberg

ACM Press

Addison-Wesley Publishing Company

© 上海科学普及出版社 1994

© 1992 年 Addison-Wesley 出版公司英文版的授权中文译本。本中文译本的出版和发行获得上述英文版的一切出版和发行权的所有者——Addison-Wesley 出版公司的许可。上海科学普及出版社拥有上述英文版在中华人民共和国（除台湾省外）的独家翻译、出版、发行权。任何人不得再对上述英文版和中文译本进行翻译、改编、出版和发行。

JS/20 / 8

## 作者名单

**Max Behensky**, Atari Games, Milpitas, CA 95035

**R.Daniel Bergeron**, Computer Science Department, University of New Hampshire, Durham, NH 03824

**Meera Blattner**, Lawrence Livermore National Laboratory, Livermore, CA 94551

**Hans-Jörg Bullinger**, Fraunhofer Institut für Arbeitswirtschaft und Organisation, Stuttgart 1, Germany

**Frederick P. Brooks, Jr.**, Department of Computer Science, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599

**Stuart K. Card**, Xerox Palo Alto Research Center, Palo Alto, CA 94304

**Kathleen D. Cebulka**, Bellcore, Piscataway, NJ 08854

**Rich Cohen**, Human Interface Laboratory, Microelectronics and Computer Consortium, Austin, TX 78759

**Roger B. Dannenberg**, School of Computer science, Carnegie Mellon University, Pittsburgh, PA 15213-3890

**Abbe Don**, Multimedia Designer, San Francisco, CA 94118

**Robert F. Farrell**, Bellcore, Piscataway, NJ 08854

**S.K.Ganapathy**, Machine Perception Research Department, AT&T Bell Laboratories, Holmdel, NJ 07733

**Rich Gossweiler**, Department of Computer Science, University of Virginia, Charlottesville, VA 22903

**Robert M.Greenberg**, San Francisco Conservatory of Music, San Francisco, CA 94122

**Georges G.Grinstein**, Computer Science Department, University of Massachusetts, Lowell, Lowell, MA 01854

**Karl-Heinz Hanne**, Fraunhofer Institut für Arbeitswirtschaft und Organisation, Stuttgart 1, Germany

**Alexander G.Hauptmann**, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-3890

**Will Hfl,** Bellcore, Morristown, NJ 07960-1910

**Robert L. Joseph,** School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-3890

**Minao Kamegai,** Lawrence Livermore National Laboratory, Livermore, CA 94551

**Srdjan Kovacevic,** College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280

**Brenda Laurel,** Telepresence Research, Inc., Palo Alto, CA 94301

**Jock D. Mackinlay,** Xerox Palo Alto Research Center, Palo Alto, CA 94304

**Tim McCandless,** U.S. West Advanced Technologies, Boulder, CO 80303

**Brian Michon,** Digital Equipment Corporation, Maynard, MA 01754-2571

**Margaret Minsky,** MIT Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139

**Michael J. Muller,** Bellcore, Piscataway, NJ 08854

**Joseph Nolthuis,** Educa Video Multimedia Design, Utrecht, The Netherlands

**Tim Oren,** Apple Computer, Inc., Advanced Technology Group, Cupertino, CA 95014

**Ming Ouh-young,** AT&T Bell Laboratories, Middletown, NJ 07748

**Andrew J. Palay,** Silicon Graphics Computer Systems, Mountain View, CA 94039

**Alan P. Parkes,** Dept. of Computing, University of Lancaster, Lancaster, United Kingdom

**Randy Pausch,** Department of Computer Science, University of Virginia, Charlottesville, VA 22903

**R. Douglas Riecken,** Integrated Computer / Communications Systems Research Department, AT&T Bell Laboratories, Holmdel, NJ 07733

**George G. Robertson,** Xerox Palo Alto Research Center, Palo Alto, CA 94304

**Alexander I. Rudnicky,** School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-3890

**Lillian S. Ruston,** Bellcore, Morristown, NJ 07962

**John G. Smith,** Bellcore, Piscataway, NJ 08854

**Stuart Smith**, Computer Science Department, University of Massachusetts, Lowell, Lowell,  
MA 01854

**Oliver Steele**, Apple Computer, Inc., Cambridge, MA 02142

**Yoichi Takebayashi**, Research and Development Center, Toshiba Corporation, Kawasaki,  
Japan

**David Weimer**, Machine Perception Research Department, AT&T Bell Laboratories,  
Holmdel, NJ 07733

**Elizabeth M. Wenzel**, Human Interface Research Branch, NASA Ames Research Center,  
Moffett Field, CA 94035

**David Wroblewski**, U.S. West Advanced Technologies, Boulder, CO 80303

**Polle T. Zellweger**, Xerox Palo Alto Research Center, Palo Alto, CA 94304

# 前言

一个为期两天，重点讨论用户接口设计中的多媒体和多模态 (multimodal) 问题和成果的研讨会在 CHI'90“计算系统中人为因素”年会上召开，此次会议由计算机联合会下属的人机交互作用特别兴趣组主办。研讨会的目标是认定这些成果，并将最终导致从结构上加强这种新技术的基础。研讨会的资料摘要收录在 1990 年 10 月 SIGCHI 简报第 22 卷第 2 册中。与会者一致认为，这次研讨会成功地达到了预期目标，所讨论的主题具有长远的意义。会议要求与会者贡献原始的资料，收集到本书中。其它一些章节则是从本次年会的多媒体论文讨论会上收集后编入的。那些虽不在本次研讨会范围之内，但我们相信会成为多媒体和多模态领域研究的中心问题的文章，是由另外一些作者提供的。

涉及的领域之宽广，加上与硬件、软件、应用以及实现有关的多方面问题交织在一起，使得多媒体和多模态用户接口的研究工作显得混沌一片。目前，多媒体接口的结构是凭借直觉、感觉以及经验来驱动的。我们仔细考虑了我们在理解这一新出现的技术的基础原理方面所要达到的目标，我们将收集到的资料分成四类。这四个类别按照以下四方面来分析接口：

- 1) 它们的模型和隐喻 (metaphor)；
- 2) 它们的构成和设计；
- 3) 它们的理论基础和实现技术；
- 4) 它们的体系结构。

本书中各个部分都和这四个方面相对应。每一部分中的大多数章节都与许多不同的问题有关，而且放在另一部分中也会很恰当。但是，每篇文章的位置都是根据它是否包含有所属类别的重要内容以及上下文的衔接来确定的。虽然各章都是每位作者努力研究的结果，但从资料的组织可以看出，多媒体和多模态接口的真正的基础是一些更为广阔的问题，包括构造模型、确定结构以及解决与新设备的结构有关的各种技术问题。

许多共同的思路贯穿于本书的各个主题。纵观模型、组成、技术和体系结构这四大部分，还可以按其它一些问题和领域将它们组织在一起。这些特定主题中感兴趣的那些主题分列于下：

## 声音 (Sound)

在这些章节中，“声音”有许多形式。非语音的声频在 Smith、Bergeron 和 Grinstein 写的第十一章“探索数据分析用的立体声和层面声音的产生”中讨论。他们的新方法通过声频解释探索数据显示上图符的特征。Blattner、Greenberg 和 Kamegai 写的第六章“用声音表达湍流：科学声频化一例”讨论如何通过起表达作用的信息以及声音消息来解释科学数据。Wenzel 写的第十五章“三维虚拟声显示”论述了为声音增加一维信息，并考虑了与之有关的问题。在 Muller、Farrell、Cebulka 和 Smith 写的第一章“随时间变化的多媒体的可用性问题”以及在 Rudnicky 和 Hauptmann 写的第十章“语音系统中的多模态交互”中讨论了语音技术。Dannenberg 和 Joseph 写的第四章“钢琴训练中的人机接口”描述了作为输入和输出媒体的音乐。Riecken 在第十八章“人机交互作用和感知过程”中讨论了“适应的听觉”。Weimer 和 Gunapathy 在第七章“使用手的运动轨迹跟踪和语音识别的交互技术”中使用了语音驱动的菜单。Takebayashi 写的第十四章“多媒体接口的理解和合成功能的综合”描述了 Bank-by-Phone 系统中所需的语音输入。另外还有几章也讨论了声音的集成，不是作为重

点，但也是系统的组成部分之一。

### 多模态输入和手势

交互技术是本书的一个重要组成部分。Card、Mackinlay 和 Robertson 写的第十三章“输入设备的设计空间”描述了输入设备的分类。Weimer 和 Ganapathy 写的第七章“使用手的运动轨迹跟踪和语音识别的交互技术”讨论了手势和语音的组合使用问题。Hanne 写的第八章“多模态通讯：文本和手势的集成”分析了许多手势驱动的接口。Pausch 和 Gossweiler 写的第九章“根据不精确的多模态输入进行与应用无关的目标选择”，以及 Rudnicky 和 Hauptmann 写的第十章“语音系统中的多模态交互”都着重叙述了对来自各种输入设备的二义性输入进行选择 and 校正的技术。Minsky、Ouh-young、Steele、Brooks 和 Behensky 写的第十六章“触觉和视觉：力显示中的一些问题”描述了一种新的输入技术，即具有触觉特征的表面的感知技术。Riecken 写的第十八章“人机交互作用和感知过程”讨论了随用户的移动而变化的自适应接口和键盘。

### 随时间变化的媒体

David Weimer 写的第一部分的引言中讲到，多媒体和多模态用户接口研究工作的每一进展都和我们对如何使用随时间变化的媒体的理解程度联系在一起。声音（非语音的和语音的）、视频、动画和手势都是随时间变化的媒体。本书的前十章包含了与随时间变化的媒体有关的资料。在后面十章中，Brian Michon 写的第二十章“高度图符化的接口”着重讨论对“流（streams）”和“块（chunks）”进行编辑时所存在的差异。Minsky、Ouh-young、Steele、Brooks 和 Behensky 在第十六章“触觉和视觉：力显示中的一些问题”中描述了对纹理（texture）的感觉。Zellweger 在第二章“活动多媒体文档的模型”中描述了使用随时间变化的注释时存在的问题。

### 应用

有一些应用需求离开了多媒体接口是不可能实现的。Dannenberg 和 Joseph 写的第四章“钢琴训练中的人机接口”中讲到，它教初学者演奏钢琴，依靠钢琴作为主要的计算机接口。依靠视频来维持学生的兴趣。交互式的视频戏剧效果（videodrama）取决于演奏者的精细交互动作。Nolthuis 在第五章“多媒体和链接现实的艺术”中描述了一个程序“Rights in the Mirror”，它使用人们之间微妙的交互作用来讲授公民权利问题。Laurel、Oren 和 Don 写的第三章“多媒体接口设计中的问题：媒体集成和接口中介”叙述怎样借助一种称为“导游”的动画特性，用历史和地理知识来激发学生的想象力，这一功能是在“美国系列：合众国历史的 CD-ROM 激光盘”中实现的。Michon 描述了 Athena Muse 科研项目中如何使用激光录象盘来建立诸如训练和模拟旅行之类的应用。Smith、Bergeron 和 Grinstein 在第十一章“探索数据分析用的立体声和层面声音的产生”中所用的 Exvis 系统是一个用来分析各种各样科学数据的工作系统。Muller、Farrell、Cebulka 和 Smith 在第一章“随时间变化的多媒体的可用性”中叙述了一个语音超级媒体系统 Hyperphone，一个教学系统 WITS 以及一个多媒体的音乐台（kiosk）。

多媒体的其它应用减轻了繁重的输入和输出任务。Hanne 描述了不用手送入东方国家的字符以及自然语言的输入，这可以简化没有经过计算机系统方面专门训练的人的工作。语

音在许多应用中很受欢迎, Takebayashi 在第十四章“多媒体接口的理解和合成功能的综合”介绍的 Bank-by-Phone 和 APS (一个自动演示系统) 中描述了语言和人机对话。Muller、Farrell、Cebulka 和 Smith 描述了一个浏览语音文档的系统, Zellweger 在第二章“活动多媒体文档的模型”中讨论了语音注释。Palay 在第十九章“未来用户接口成分的‘操作系统’”中描述了 Andrew 工具包。Michon 在第二十章“高度图符化的接口”中说, Athena Muse 具有各种视频编辑能力。该章描述了使用一个 X 窗口多媒体结构包的 Athena Visual Computing Group 所完成的 8 个科研项目。

## 体系结构的原理和理论

Hill、Wroblewski、McCandless 和 Cohen 在第十七章“多模态和多媒体接口的体系结构品质和原则”中讨论了多媒体和多模态接口的 5 个重要品质, 以及“统一访问原则 (Principle of Uniform Access)”。Parkes 在第十二章中讨论了“可管理的媒体间编码的推理适当性”, 也就是将一种媒体里的表达转换成另一种媒体里的表达。由于表达结构在某些方面可能类似, 在其它方面又可能有所不同, 由此带来的问题会将轻信的用户引入歧途。这一转换问题在图形用户接口 (GUI) 中显得特别重要, 因为在 GUI 中有许许多多的应用是以图形形式展示的。在 Card、Mackinlay 和 Robertson 写的第十三章“输入设备的设计空间”中描述了输入设备的分类。Pausch 和 Gossweiler 写的第九章“根据不精确的多模态输入进行与应用无关的目标选择”讨论了一个用于选择目标的数学模型。音乐理论是 Blattner、Greenberg 和 Kamegai 写的第六章“用声音表达湍流: 科学声频化一例”的核心。

## 作为剧本的接口

有两章与作为剧本的接口设计有关。Laurel、Oren 和 Don 写的第三章“多媒体接口设计中的问题: 媒体集成和接口中介”描述了使用剧本特性来集成各种媒体。这些特性集中于该章的说明部分。Nolthuis 在“多媒体和链接现实的艺术”一章中, 借用 Bertolt Brecht 创作的练习剧中所需的方法, 用戏剧中的方法来讲授公民权利问题。

## 智能多媒体接口

Riecken 写的第十八章“人机交互作用和感知过程”的重点是与用户需求相适配的系统。Dannenberg 和 Joseph 在 Piano Tutor 程序中要求对学生的需求作智能的交互和分析。许多章节中都涉及到自然语言方面的问题。Takebayashi 在第十四章“多媒体接口的理解和合成功能的综合”中推荐了一些基于合成和理解功能的媒体功能, 使得能在多媒体/多模态接口中增加智能性。Hanne 在第八章“多模态通讯: 文本和手势的集成”在考虑了自然语言输入和手势相组合后所起的作用。Rudnicky 和 Hauptmann 在第十章“语音系统中的多模态交互”中讨论了自然语言系统中的出错纠正问题。Muller、Farrell、Cebulka 和 Smith 在“随时间变化的多媒体的可用性问题”一章中也描述了自然语言输入的使用。Kovacevic 在第二十一章“人机对话的组合模型”中描述了一个自然语言接口。

## 多媒体操作系统和对话管理

Palay 在第十九章“未来用户接口成分的‘操作系统’”中描述了 Andrew Toolkit (工具包), 这是一个依赖于相互嵌入的应用程序成分的操作系统。Kovacevic 在第二十一章“人机对话的组合模型”中考虑了构造人机对话模型的问题, 并描述了一个协调这一模型中多方面问题的用户接口管理系统。Takebayashi 在第十四章“理解和合成功能的综合”中为 Bank-by-Phone 提出了对话管理的要求。语音和差错校正的交织问题是 Rudnicky 和 Hauptmann 写的第十章“语音系统中的多模态交互”的主题, 而且是 Carnegie Mellon 大学的一个更大的科研项目 (称为 MINDS, 研究对话的管理) 的组成部分。对话管理是 Nolthuis 写的第五章“多媒体和链接现实的艺术”中的关键要素。Muller、Farrell、Cebulka 和 Smith 在第一章“随时间变化的多媒体的可用性问题”中介绍了在 HyperPhone 科研项目中使用的对话管理模型。关于对话的见解集中在 Dannenberg 和 Joseph 写的第四章“钢琴训练中的人机接口”中。

本书是面向研究和开发方面的问题的, 因此, 我们在书中回避了实现交互式多媒体接口时的实践问题。仅在谈到新的模型或技术的上下文中才讨论有关的应用。多媒体接口设计方面的进展是非常快的。新的多媒体技术随时准备推向公众市场。用户已经表明, 他们对于计算机工业所能给予他们的“新玩意”的需求是贪得无厌的。现在需要提出的问题是认定一些基础性的问题, 我们希望, 对于这些问题的答案已经出现了一些光芒。

## 致谢

本书是 ACM 人机交互作用特别兴趣组 (SIGCHI) 主办的多媒体和多模态用户接口设计 CHI'90 研讨会的副产品。如果没有 SIGCHI 提供的特别协助, 就不可能写成这本书。此外, Lawrence Livermore 国家实验室计算室副主任 Robert Borchers 和计算室助理副主任 John Ranelletti 提供了经济资助和鼓励。Addison-Wesley 的 Peter Gordon 和 Helen Goldstein 以及 ACM 出版社的 Nhora Cortes-Comerer 在出版工作方面给了我们很多指导。Peter Gordon 早期对这一项目的支持, 包括参加研讨会一直到他对本书的组织 and 风格提出的许多建议, 都为我们提供了完成预期目标的方法。最后, 我们还要感谢我们的书评作者, Virginia 工艺学院和州立大学的 Edward Fox, 以及加利福尼亚州洛杉矶的 Philips InteractiveMedia 的 Blacksburg、Virginia 和 Tyler Blake, 他们对本书所作的许多有益的评注大大提高了本书的质量。

Meera M. Blattner 和 Roger B. Dannenberg

# 目 录

|                                                  |    |
|--------------------------------------------------|----|
| 引言 多媒体接口的趋势 .....                                | 1  |
| 0.1 为什么需要多媒体系统? .....                            | 2  |
| 0.2 多媒体的目标: 更好的通讯手段 .....                        | 2  |
| 0.2.1 计算机对计算机的通讯 .....                           | 2  |
| 0.2.2 人对人的通讯 .....                               | 3  |
| 0.2.3 计算机与人之间的通讯 .....                           | 4  |
| 0.3 术语 .....                                     | 5  |
| 0.3.1 多媒体 .....                                  | 5  |
| 0.3.2 编码 (encodings) 和表达 (representations) ..... | 5  |
| 0.3.3 方式 (mode) 和模态 (modality) .....             | 6  |
| 0.3.4 多通道 .....                                  | 6  |
| 0.4 多媒体的现状 .....                                 | 6  |
| 第一部分 模型、隐喻和范例 .....                              | 9  |
| 第一部分的引言 .....                                    | 11 |
| 1.1 引言 .....                                     | 11 |
| 1.2 单一的和混合的媒体格式之特性 .....                         | 12 |
| 1.3 应用领域 .....                                   | 12 |
| 1.4 信息和用户的构造 .....                               | 12 |
| 第一章 随时间变化的多媒体的可用性问题 .....                        | 14 |
| 1.1 引言 .....                                     | 14 |
| 1.2 多媒体的科研项目 .....                               | 14 |
| 1.2.1 语音超级媒体 HyperPhone .....                    | 14 |
| 1.2.2 多媒体音乐台 (kiosk) .....                       | 15 |
| 1.2.3 教学用超级媒体 WITS .....                         | 16 |
| 1.2.4 用户的任务 .....                                | 17 |
| 1.3 随时间变化的媒体之特性 .....                            | 17 |
| 1.3.1 通讯模型 .....                                 | 17 |
| 1.3.2 灵活的交互动作和速度调整 (pacing) .....                | 18 |
| 1.3.3 上下文敏感性问题 .....                             | 18 |
| 1.4 多媒体应用环境的体系结构 .....                           | 19 |
| 1.4.1 Hyper Phone .....                          | 19 |
| 1.4.2 WITS .....                                 | 21 |
| 1.4.3 Kiosk .....                                | 24 |
| 1.5 可用性问题 .....                                  | 24 |
| 1.5.1 颗粒性 (granularity) .....                    | 25 |
| 1.5.2 复杂性 (complexity) .....                     | 25 |

|                                              |           |
|----------------------------------------------|-----------|
| 1.5.3 执行链路 .....                             | 26        |
| 1.5.4 领会用户需要 .....                           | 29        |
| 1.6 对随时间变化的接口的改进 .....                       | 29        |
| 1.6.1 语音的直接处理 .....                          | 30        |
| 1.6.2 创作 .....                               | 30        |
| 1.6.3 为 Hyper Phone 增加智能 (见注 7) .....        | 31        |
| 1.7 有关的工作 .....                              | 35        |
| 1.8 小结 .....                                 | 36        |
| <b>第二章 活动多媒体文档的模型 .....</b>                  | <b>38</b> |
| 2.1 引言 .....                                 | 38        |
| 2.2 文档结构 .....                               | 38        |
| 2.3 EtherPhone 系统的多媒体能力 .....                | 39        |
| 2.4 多媒体注释 .....                              | 40        |
| 2.4.1 音频注释和编辑 .....                          | 40        |
| 2.4.2 视频注释和编辑 .....                          | 42        |
| 2.4.3 注释模型方面的经验 .....                        | 42        |
| 2.5 路径模型 .....                               | 43        |
| 2.5.1 原本型文档系统 .....                          | 43        |
| 2.5.2 路径模型方面的经验 .....                        | 46        |
| 2.6 小结以及今后的工作 .....                          | 47        |
| <b>第三章 多媒体接口设计中的问题: 媒体集成和接口中介* .....</b>     | <b>48</b> |
| 3.1 引言: 一个正在研制中的多媒体接口 .....                  | 48        |
| 3.1.1 主题 .....                               | 48        |
| 3.1.2 背景 .....                               | 48        |
| 3.1.3 术语 .....                               | 49        |
| 3.2 媒体集成 .....                               | 49        |
| 3.2.1 媒体“居住区 (Ghettos)” .....                | 49        |
| 3.2.2 媒体偏见 (Biases) .....                    | 50        |
| 3.2.3 跨媒体的链路 .....                           | 52        |
| 3.3 指南 .....                                 | 53        |
| 3.3.1 超级文本模型方面的问题 .....                      | 53        |
| 3.3.2 指南的作用: 通过着眼点 (point of view) 来集成 ..... | 53        |
| 3.3.3 特征化 .....                              | 54        |
| 3.3.4 “指南”如何才能支持用户 .....                     | 54        |
| 3.3.5 期望的目标 .....                            | 56        |
| <b>第四章 钢琴训练中的人机接口 .....</b>                  | <b>57</b> |
| 4.1 引言 .....                                 | 57        |
| 4.2 系统综述 .....                               | 57        |

|                                       |           |
|---------------------------------------|-----------|
| 4.3 内容和媒体的分离 .....                    | 59        |
| 4.4 演示系统 .....                        | 59        |
| 4.4.1 其它特性 .....                      | 61        |
| 4.4.2 演示构造程序的优点 .....                 | 61        |
| 4.5 对话 .....                          | 62        |
| 4.6 学生模型和超级媒体 .....                   | 63        |
| 4.7 小结 .....                          | 65        |
| <b>第五章 多媒体与链接现实的艺术 .....</b>          | <b>67</b> |
| 5.1 引言 .....                          | 67        |
| 5.2 程序的目标 .....                       | 67        |
| 5.3 程序的使用者 .....                      | 68        |
| 5.4 和使用者之间的交互作用 .....                 | 68        |
| 5.5 屏幕设计 .....                        | 68        |
| 5.6 戏剧的应用 .....                       | 70        |
| 5.7 教学设计 .....                        | 71        |
| 5.8 小结 .....                          | 72        |
| <b>第六章 用声音表达湍流: 科学声频化一例 .....</b>     | <b>73</b> |
| 6.1 引言和背景 .....                       | 73        |
| 6.2 数据表达与消息的对比 .....                  | 74        |
| 6.3 流体中的湍流 .....                      | 74        |
| 6.4 在湍流研究中利用声音 .....                  | 75        |
| 6.4.1 表达 .....                        | 75        |
| 6.4.2 声音消息 .....                      | 77        |
| 6.4.3 先例和感知能力 .....                   | 77        |
| 6.5 湍流的声音参数 .....                     | 78        |
| 6.5.1 描述流体特征的声音 .....                 | 78        |
| 6.5.2 描述运动特征的声音 .....                 | 79        |
| 6.5.3 描述涡旋特征的声音 .....                 | 80        |
| 6.5.4 能量的消散 .....                     | 80        |
| 6.6 声音消息 .....                        | 81        |
| 6.6.1 声符: 对表达的扩展和增强 .....             | 81        |
| 6.6.2 声符: 提供数据表达方式所不能提供的信息 .....      | 82        |
| 6.6.3 声符和起表达作用的声音: 设计指南 .....         | 82        |
| 6.7 先例和可听度: 同时使用表达用的声音和声音消息时的模型 ..... | 82        |
| 6.7.1 结构 .....                        | 83        |
| 6.7.2 我们的系统所依据的音乐发展历史 .....           | 83        |
| <b>第二部分 构成和组合 .....</b>               | <b>85</b> |

|                                     |     |
|-------------------------------------|-----|
| 第二部分的引言 .....                       | 87  |
| 第七章 使用手的运动轨迹跟踪和语音识别的交互技术 .....      | 89  |
| 7.1 引言 .....                        | 89  |
| 7.1.1 进行中的协合工程工作 .....              | 89  |
| 7.1.2 在协合接口中的混合各种模态 .....           | 91  |
| 7.2 多种模态的混合 .....                   | 92  |
| 7.2.1 人与人对话的一些例子 .....              | 92  |
| 7.2.2 三维轨迹跟踪 .....                  | 92  |
| 7.2.3 增加语音识别 .....                  | 93  |
| 7.3 用于计算机辅助设计的输入技巧 .....            | 94  |
| 7.3.1 使用手的三种方法 .....                | 94  |
| 7.3.2 一个实验性的手势加语音的系统 .....          | 97  |
| 7.3.3 三维设计的某些考虑 .....               | 97  |
| 7.4 小结 .....                        | 100 |
| 第八章 多模态通讯: 文本和手势的集成 .....           | 102 |
| 8.1 人机接口中的限制 .....                  | 102 |
| 8.2 人机交互作用风格 .....                  | 102 |
| 8.2.1 自然语言交互过程对 HCI 来说“自然”吗? .....  | 103 |
| 8.2.2 直接处理“直接在屏幕上吗?” .....          | 103 |
| 8.2.3 形式语言 .....                    | 104 |
| 8.3 手势: 只是自然语言与直接处理的组合还是更进一步? ..... | 104 |
| 8.3.1 HCI 中为什么要用手势? .....           | 104 |
| 8.3.2 自然语言和指点的手势 .....              | 105 |
| 8.4 在组合的 HCI 系统中的手势 .....           | 105 |
| 8.5 手势接口中的应用 .....                  | 106 |
| 8.5.1 计算机盲人用户的手势接口 .....            | 107 |
| 8.5.2 手势驱动的中文和日文文本编辑程序 .....        | 107 |
| 8.5.3 手势和手写印刷体字母 .....              | 107 |
| 8.5.4 手势和校正符号 .....                 | 107 |
| 8.6 将来的方案: 语义学的鼠标器/笔记本 .....        | 109 |
| 第九章 根据不精确的多模态输入进行与应用无关的目标选择 .....   | 110 |
| 9.1 引言 .....                        | 110 |
| 9.2 模型 .....                        | 111 |
| 9.3 实现 .....                        | 113 |
| 9.4 小结 .....                        | 114 |
| 第十章 语音系统中的多模态交互 .....               | 115 |
| 10.1 引言 .....                       | 115 |

|                                           |            |
|-------------------------------------------|------------|
| 10.1.1 背景 .....                           | 115        |
| 10.1.2 语音识别系统 .....                       | 116        |
| 10.1.3 语音的语言系统所提出的特殊问题 .....              | 117        |
| 10.2 语音语言系统的设计原则 .....                    | 119        |
| 10.2.1 用户可塑性 .....                        | 119        |
| 10.2.2 交互协议的风格 .....                      | 120        |
| 10.2.3 纠错用的多种模态 .....                     | 124        |
| 10.2.4 响应时间 .....                         | 127        |
| 10.2.5 任务特定的对话结构 .....                    | 129        |
| 10.2.6 多模态交互作用 .....                      | 130        |
| 10.3 小结 .....                             | 131        |
| <b>第十一章 探索数据分析用的立体声和层面声音的产生 (注) .....</b> | <b>133</b> |
| 11.1 引言 .....                             | 133        |
| 11.2 数据的图解 (iconographic) 显示 .....        | 133        |
| 11.3 用声音表达数据 .....                        | 136        |
| 11.4 有关的研究工作 .....                        | 136        |
| 11.5 立体声的声音显示 .....                       | 137        |
| 11.6 层面声音的产生 .....                        | 138        |
| 11.7 摘要 .....                             | 139        |
| <b>第三部分 前提和使之实现的技术 .....</b>              | <b>141</b> |
| <b>第三部分的引言 .....</b>                      | <b>143</b> |
| <b>第十二章 可管理的媒体间编码 (MIME) 的推理适当性 .....</b> | <b>145</b> |
| 12.1 引言 .....                             | 145        |
| 12.2 表达、用户计算和媒体 .....                     | 146        |
| 12.3 可管理的媒体间编码 (MIME) .....               | 147        |
| 12.3.1 现有系统中的某些 MIME .....                | 148        |
| 12.3.2 怎样才能断定用户已作了连接? .....               | 149        |
| 12.4 利用 MIME 的推理适当性 .....                 | 150        |
| 12.4.1 任务 .....                           | 150        |
| 12.4.2 系统 .....                           | 151        |
| 12.4.3 研究 .....                           | 154        |
| 12.4.4 问题和答案 .....                        | 156        |
| 12.5 讨论 .....                             | 161        |
| 12.6 小结 .....                             | 163        |
| <b>第十三章 输入设备的设计空间* .....</b>              | <b>167</b> |
| 13.1 引言 .....                             | 167        |

|                                    |     |
|------------------------------------|-----|
| 13.1.1 工具箱 (Toolkits)              | 167 |
| 13.1.2 分类学 (Taxonomies)            | 167 |
| 13.1.3 性能研究                        | 168 |
| 13.2 分析性的设计框架                      | 168 |
| 13.3 生成设计空间                        | 168 |
| 13.3.1 基本移动词汇表                     | 169 |
| 13.3.2 组合操作符                       | 170 |
| 13.3.3 输入设备的设计空间                   | 171 |
| 13.4 设计空间中的测试点                     | 172 |
| 13.4.1 表达性                         | 172 |
| 13.4.2 有效性                         | 173 |
| 13.4.3 足迹                          | 174 |
| 13.4.4 带宽                          | 174 |
| 13.4.5 实例: 用鼠标器和头戴式鼠标器显示三维信息环境中的选择 | 176 |
| 13.5 结束语                           | 177 |
| <b>第十四章 多媒体接口的理解和合成功能的综合</b>       | 179 |
| 14.1 概述                            | 179 |
| 14.2 理解和合成的映射模型                    | 180 |
| 14.3 使用理解和合成功能的多媒体应用系统             | 182 |
| 14.3.1 自动演示系统                      | 182 |
| 14.3.2 使用与说话者无关的语音识别的电话银行系统        | 185 |
| 14.4 讨论                            | 193 |
| <b>第十五章 三维虚拟声显示</b>                | 195 |
| 15.1 引言                            | 195 |
| 15.2 为什么要用虚拟声显示                    | 195 |
| 15.3 三维虚拟声显示的经历                    | 198 |
| 15.3.1 心理声学的经历                     | 198 |
| 15.3.2 实现的方法                       | 200 |
| 15.4 NASA AMES 的三维听觉显示项目           | 204 |
| 15.4.1 实时系统: Convolvotron          | 205 |
| 15.5 合成技术的心理物理学验证                  | 207 |
| 15.5.1 验证使用个性化 HRTF 的静止声源          | 208 |
| 15.5.2 定位性能的声学要素                   | 210 |
| 15.5.3 无经验的收听者和非个性化的 HRTF          | 213 |
| 15.6 改进虚拟声显示: 问题范围和研究课题            | 216 |
| <b>第十六章 触觉和视觉: 力显示中的一些问题*</b>      | 219 |
| 16.1 “砂纸”系统                        | 219 |
| 16.1.1 为什么要产生纹理                    | 220 |