

面向儿童的多通道交互系统*

李杰, 田丰, 王维信, 戴国忠

(中国科学院 软件研究所 人机交互技术与智能信息处理实验室, 北京 100080)

E-mail: lijie@iel.iscas.ac.cn

http://iel.iscas.ac.cn

摘要: 设计和实现了一个基于笔和语音的面向儿童的多通道三维交互系统.系统中包含了基于笔和语音的交互信息整合框架,用来整合儿童输入的笔和语音信息.同时,系统中定义了一些基于笔和语音的交互技术,可以支持孩子们以自然的方式,通过笔和语音同系统进行交互.用笔来勾画三维的场景和小动物等实体,同时用笔和语音同场景和场景中的实体进行一定的交互.

关键词: 多通道;儿童;手势;语音检测;交互原语;交互任务

中图法分类号: TP391 **文献标识码:** A

多通道技术在许多面向儿童的交互系统中已经得到广泛的应用.Takeo Igarashi 设计并实现了一个基于勾画的3D绘画模型Teddy^[1].利用Teddy模型,用户采用自由随意的风格便可进行3D建模.Teddy的交互方式主要采用笔和手势.Teddy并非一个精确的设计工具,它仅能生成粗略的3D模型.但是它能够及时建模,适合儿童与非专业人士使用.MIT计算机科学实验室CGG(Computer Graphics Group)小组致力于设计一个用以组成和展示透视图景的投影绘画系统^[2].这个系统采用一种2D的投影表现方法替代传统的3D描述,使绘制投影图景变得如同传统2D绘图一样轻松.用户可以轻松地采用笔结合手势操作3D变换^[3].投影绘画系统的交互方式是笔.MIT媒体实验室设计并开发了一种新颖的锻炼儿童会话能力的界面系统Doll Talk^[4].Doll Talk通过捕捉儿童的手势、语音信息来模拟语音识别,通过改变音调的方式将会话内容反馈给儿童,并引导儿童改善自己的叙述.Doll Talk良好的语音界面与交互对增强孩子的表达能力效果明显.国内最近也有很多研究机构与公司制作了一些类似的面向儿童的商业软件,允许使用者利用鼠标或手写笔等2D输入设备在平面上进行绘图操作.但这类软件或系统大多不具备3D功能以及基于语音、语调等自然交互方式进行交互的能力.

但是,对儿童这一特殊用户群体而言,使用常规多通道交互系统仍然显得困难而且枯燥乏味.其中的关键问题是儿童的成熟度以及认知度都不足以使之精确定义并描述基于笔、手势、语音等多种交互设备的交互操作,而且对应用的表现而言,针对实体和场景的指定及描述过于精确对于儿童用户是不必要的负担.

1 系统概述

本文设计并实现了一个面向儿童的多通道交互系统“神笔马良”.它是一个针对儿童设计开发的绘画应用.在本系统中,儿童用户可以使用笔描绘背景,以随意自由的勾画形式描绘三维小动物实体;在使用过程中,可以结合语音语调、手势等自发的信息辅助绘图;同时可以使用笔和手势对绘画对象进行三维交互.系统应用面向儿童,在设计中充分考虑到儿童用户与计算机交互时相对随意、不精确并且追求自然的特点,交互界面简单明

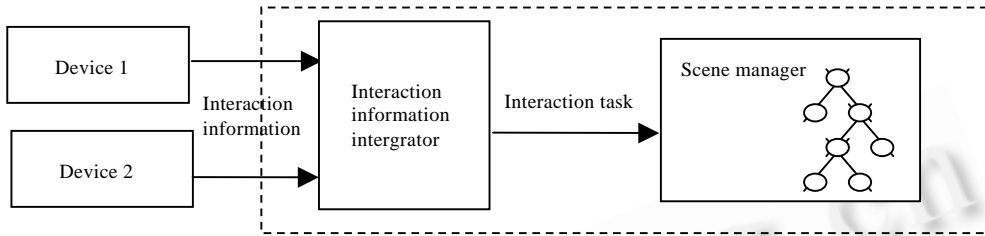
* 收稿日期: 2002-03-05; 修改日期: 2002-06-05

基金项目: 国家高技术研究发展计划资助项目(2001AA114170)

作者简介: 李杰(1977-),男,湖北荆门人,博士生,主要研究领域为人机交互技术,笔式计算;田丰(1976-),男,陕西绥德人,博士生,主要研究领域为人机交互技术,虚拟现实;王维信(1970-),男,吉林农安人,博士,主要研究领域为计算机图形学,人机交互技术;戴国忠(1944-),男,江苏无锡人,研究员,博士生导师,主要研究领域为人机交互技术,计算机图形学.

了.在交互过程中,以笔为主要工具,结合手势和音调完成各种输入和操作,并提供语音和音乐等提示作为信息反馈,引导用户成功地完成交互任务.

系统采用基于自然交互方式的多通道用户界面模型进行设计^[5],通过整合来自不同通道的自然方式输入来完成界面的操作.其层次框架如图 1 所示.



设备 1, 设备 2, 交互信息, 交互信息整合器, 交互任务, 场景管理器。

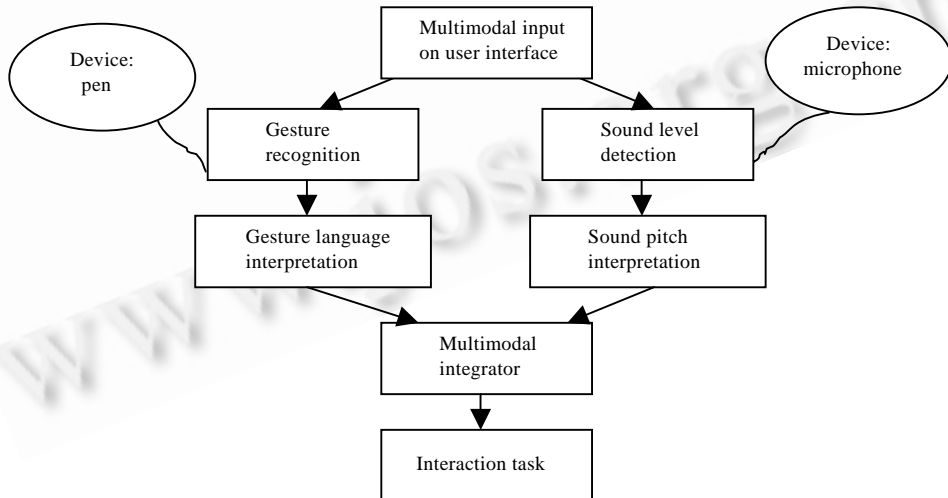
Fig.1 Framework of Toolkit

图 1 神笔马良系统框架

此系统主要由交互信息整合器和场景管理器两大部分组成.交互信息整合器负责整合来自不同交互设备的信息,并将其转化为特定的交互任务.本系统中主要的交互通道包括手写笔和语音,交互设备包括手写笔、语音设备.系统支持交互设备的扩充,交互设备整合器可以根据实际需要来增加相应的交互设备.场景管理器负责管理三维系统中的对象.它接收到交互信息整合器发来的交互任务以后,按照不同任务的具体要求更改场景中对象的属性,进行相应的语法和语义反馈.

按照系统模型的层次划分,用户可以完成不同通道与不同功能之间的映射.由于实现了交互任务与通道的无关性、交互操作与通道的无关性,因此,系统本身具有高度的可扩充性.采用分层的结构,多通道整合在交互任务执行之前就已完成,并且系统具有良好的语义反馈能力.

依靠声强、音调检测以及基于笔的手势、指点等输入手段,系统主要整合了语音和笔两个通道的输入信息.系统的具体实现和处理流程结构如图 2 所示,归纳起来可分为 5 个步骤:(1) 手势信号识别;(2) 音调信号检测;(3) 音调强度解释;(4) 手势语言解释;(5) 最终多通道统一执行解释.



用户界面多通道输入, 手势识别, 语音检测, 手势语言翻译, 声调翻译, 多通道整合器, 交互任务, 设备:笔, 设备:麦克风.

Fig.2 Processing flow for handling information of parallel speech and gesture input

图 2 手势和语音信息并行处理流程

为了处理来自不同通道的交互信息,系统采用了一种联合解释策略^[5].这种策略基于对音调强度输入以及笔输入信号的并行处理.在处理过程中,系统在识别手势信息的同时检测音调强度.结合应用运行时的上下文,

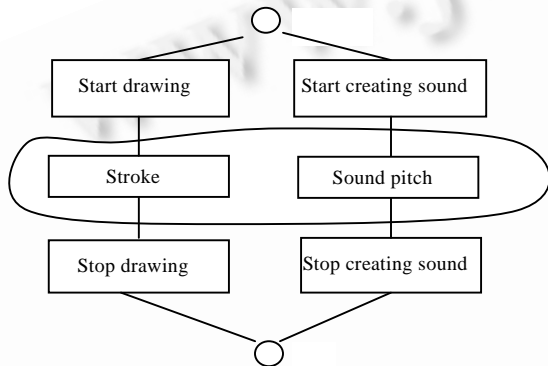
系统针对输入信息分别并行给出手势语言解释以及音调强度解释^[6]。按照事先约定的优先级(优先级由统计得到),系统自动列出手势解释和音调强度解释列表。系统根据手势与语音的优先级排列得到最终的多通道统一执行解释。经过这一过程,用户与系统完成一次交互。

采用这种基于并行识别的联合解释策略,系统在处理用户的交互信息时不依赖于某一特定通道,可以提高系统的效率及运行鲁棒性,当任何一个通道因故不能使用或者有条件通道扩展添加时均可保证系统正常运行。而基于统计得到的优先排列方法使得消除多通道二义性(mutual disambiguation)的能力大大增强^[7]。

2 多通道整合

多通道整合包括交互设备、交互方式的交融以及多个通道在意义上的传达和协作。对多通道界面而言,多通道整合显然是至关重要的。只有通过多个通道的协作,多通道的优势才能得到充分的体现。

本系统处理来自语音和手写笔两个通道的信息。多通道整合模型采用面向任务的设计思想进行设计。整合模型强调交互信息的充分性,而不要求额外的精确性^[5]。利用交互任务的结构性、时间相关性、上下文等要素来达到信息的融合。该模型以格(lattice)这种代数结构为基础。来自两个通道的输入在时间上的关系是一种偏序关系,所有事件形成一个大的偏序集。在这个偏序集中,来自同一设备的输入事件按照时间戳具有全序关系^[5],可以认为它保证了用户相应动作的顺序。



起笔, 笔画, 停笔, 声音产生, 声音强度, 声音停止。

Fig.3 Process of creating a stroke

图3 笔画产生过程

图3显示出了与一个任务——“画一个笔画(drawing a stroke)”相应的段在整个多通道输入信息流中的地位。直线段描述了针对某个通道的全序关系,曲线围绕的部分表明信息融合后判断的结果。结的上端和下端描述了其前后的任务。在勾画一个笔画的过程中,音调强度信息作为参数之一——辅助手势信息作图,当手势信息很充分时,应该开始执行任务,而不必等到信息足够精确为止。

整合主要围绕任务概念来进行。首先识别当前任务,确定任务和参数的结构,将这个结构框架应用于多通道输入的解释。在系统内,模块之间通过消息进行通信。在任务执行过程中,手势指定任务标识,音调控制强度信息或交互法则,体现多通道的互补和信息融合。信息融合算法是基于时间驱动的。

3 交互技术设计

系统采用纸和笔的隐喻(pen-based metaphor),结合音调强度来体现多通道交互的思想。纸笔隐喻由现实生活中最常用、最自然的绘画方式演变而来,因此,以手写笔为交互通道,用手势进行绘画操作是儿童最能直观接受的交互方式。语音是人们最常用的交互手段之一,利用语音可以简单地产生指令。更重要的是,语音的许多特征(如高低、快慢程度等)常常与当事人的情绪和感受密切相关。这一特征对于儿童尤其重要。采用语音交互极大地增强了儿童使用系统的兴趣。在交互过程中,以手写笔和语音设备作为输入设备,以二维平面上的三维显示作为输出,交互过程中辅以信息反馈。面向儿童的自然、简洁和无拘束是交互设计的主要目的。

系统中定义了所用到的交互原语。交互原语 IP(interaction primitive)代表了用户与计算机之间的词法级交互(输入、输出)。它是来自不同通道的独立的、最小的、不可分割的操作,这些原子操作在一定的应用上下文中有着特定的交互意义。交互原语可定义为一个四元组^[5]:

⟨交互动作,数据表示,使用通道,时间标签⟩。

交互动作表明用户与计算机之间交互的具体执行,通常与交互任务相对应。数据表示描述动作的各种属性。

使用通道指示交互动作所处的通道,用于多通道的整合.时间标签表明动作执行的时间戳,同一通道的交互在时间上的顺序关系是通道内语法分析的基础,而不同通道交互动作时间上的接近性是建立跨通道关系的基本依据之一.

在本系统中,我们将交互任务分为场景级任务和实体级任务两大类,分别对应于场景和场景中的三维实体接受交互原语所产生的交互任务.场景级任务包括场景浏览、实体选择、场景数据访问;实体级任务包括实体操纵和实体数据访问^[5].任务整合也依此分为场景级任务的整合和实体级任务的整合两大部分.通过原语分发装置进行分发判断,原语控制装置用来控制任务的整合,同时进行相应的词法反馈.

以绘图过程中的实体旋转为例,如图 4 所示.

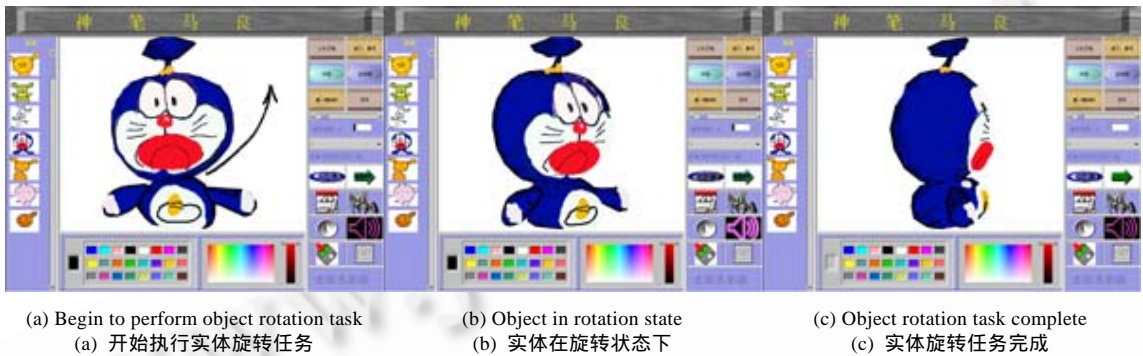


Fig.4 Process of rotating a object

图 4 实体旋转过程

这个实体级任务对应于两个交互原语——Locate 和 Quantify. Locate 以笔为交互通道,用落笔、手势作为交互动作,用坐标和方向上的偏移作为动作属性,并附以时间戳. Quantify 以语音为交互通道,以用户发声作为交互动作,以声音强度作为动作属性,并附以时间戳.下面我们分别加以具体描述.

Locate:

$\langle \text{tap/gesture, mode}/\Delta x/\Delta y/\Delta z/\Delta \alpha/\Delta \beta/\Delta \gamma, \text{pen-based, time-stamp} \rangle$

Quantify:

$\langle \text{make sound, vector/sound pitch, voice, time-stamp} \rangle$

交互任务整合器整合原语 Locate, Quantify, 使之可以执行一个完整的交互动作. 整合的主要依据是四元组中的时间戳.

系统中的交互技术包括两大类集合: 笔式交互技术集合 (pen_based) 和语音交互技术集合 (voice). 笔式交互技术有点击、持续点击、顿笔、手势输入、缩放等. 语音交互技术包括音调检测、音乐反馈、语音反馈等.

Pen_based = { tap, continuously tap, hold, gesture input, zoom };

Voice = { speech detection, music feedback, speech feedback }.

系统首先捕获到笔交互信息, 根据约定和规则推断用户所要执行的交互任务. 根据旋转手势 (方向箭头) 确定当前交互任务为“对象旋转”(如图 4(a)所示). 此时, 根据时间戳捕获对应的音调强度信息, 系统将音调强度信息作为交互任务参数. 结合当前交互任务 (对象旋转) 以及应用上下文, 系统可以确定交互任务参数的类型为“旋转速度”, 参数值根据事先约定的音调强度决定. 这里, 旋转速度与音调强度成正比. 另一个交互任务参数“旋转角度”由旋转手势得到, 旋转角度与方向箭头线段的长度成正比. 我们定义描述一个交互任务的三元组:

$\langle \text{交互对象, 任务类型, 参数列表} \rangle,$

并由此得到“对象旋转”交互任务三元组:

$\langle \text{entity, rotation, \{ velocity, angle \}} \rangle.$

系统根据交互任务完备信息开始执行旋转操作, 旋转速度由参数 velocity 决定 (如图 4(b)所示). 当旋转角度满足参数 angle 时, 停止旋转, 当前交互任务完成 (如图 4(c)所示).

此系统以音调强度作为语音信号的信息源.采用音调取代语音识别基于以下几点考虑:首先,由于儿童语音的特殊性,目前并不存在特别可行的准确识别儿童语音的方法;此外,约束性并不十分强烈的音调更符合儿童特性,并能与儿童绘画时的情绪相结合.例如,系统通常认为较高的音调输入显示了用户较强烈的表现情绪,如更快的绘画速度或更强的形变程度.

系统采用音乐和语音形式结合系统输出给出应用反馈,以引导儿童正确地完成一个完整的操作.在交互过程中,结合上下文尽可能地防止和纠正用户出现的操作错误.对于手势输入与反馈,必要的容错和错误抛弃可以激发儿童使用系统的兴趣和信心.

4 讨论

在面向儿童的多通道系统设计时,如何使界面更加简洁、更易掌握、更能激发儿童的兴趣,这是我们所要解决的大问题.本文实现了一个基于笔和语音交互的儿童绘画系统.儿童用户可以使用本系统,采用简单的方式绘画三维小动物实体,并可对绘画对象进行旋转、场景漫游等三维交互.10名5~10岁的儿童参与了系统的测试.测试结果表明,儿童不需要太多的指导和讲解即能很快掌握系统的交互方法,并对掌握系统的功能和使用该系统表现出了很大的信心和兴趣,语音与音乐提示起到了很好的引导和纠错作用.

当前有很多机构和实体也在进行儿童界面的研究与开发,其中也存在一些问题.比如,有些系统交互方式单一,无法捕获儿童释放的全部信息^[1];有些交互技术对儿童并不适用,儿童难以掌握.与同类系统相比,本系统在一定程度上克服了当前普遍存在的儿童相对难以使用的问题.儿童通过笔和语音的结合可以快速、便捷地与系统进行交互,基于简单笔画的勾画方式使得在二维画板上设计和展现三维小动物实体成为可能.在本系统的基础上,若增强智能的交互场景判断,并且融入语音中自然语言的识别与处理,则可以构造功能更为复杂和强大的儿童系统.这也是我们下一阶段所要考虑的问题.

致谢 中国科学院软件研究所人机交互技术与智能信息处理实验室的栾尚敏博士后对本文的完成提出了很多有益的建议,在此表示感谢.

References:

- [1] Igarashi, Takeo, Matsuoka, Satoshi, Tanaka, Hidehiko. Teddy: a sketching interface for 3D freeform design. In: Waggenspack, W., ed. Computer Graphics Proceedings, Annual Conference Series, SIGGRAPH'99. New York: ACM Press, 1999. 409~416.
- [2] Tolba, O., Dorsey, J., McMillan, L. A projective drawing system. In: Hughes, J.F., Séquin, C.H., eds. Proceedings of the 2001 ACM Symposium on Interactive 3D Graphics. New York: ACM Press, 2001. 25~34.
- [3] Tolba, O., Dorsey, J., McMillan, L. Sketching with projective 2D strokes. In: van der Zanden, B., Marks, J., eds. Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology. New York: ACM Press, 1999. 149~157.
- [4] Vaucelle, C., Jehan, T. Dolltalk: a computational toy to enhance children's creativity. In: Bursleson, W., Selker, T., eds. ACM CHI 2002 Conference Proceedings. New York: ACM Press, 2002. 776~777.
- [5] Dong, Shi-hai, Wang, Jian, Dai, Guo-zhong. Human-Computer Interaction and Multimodal User Interface. 1st ed., Beijing: Science Press, 1999 (in Chinese).
- [6] Oviatt, S. Multimodal system processing in mobile environments. In: Ackerman, M., Edwards, K., eds. Proceedings of the 13th Annual ACM Symposium on User Interface Software Technology (UIST 2000). New York: ACM Press, 2000. 21~30.
- [7] Cohen, P.R., McGee, D., Clow, J. The efficiency of multimodal interaction for a map-based task. In: Nirenburg, S., ed. Proceedings of the Applied Natural Language Processing Conference (ANLP 2000). New York: ACM Press, 2000. 331~338.

附中文参考文献:

- [5] 董士海,王坚,戴国忠.人机交互和多通道用户界面.北京:科学出版社,1999.

A Multimodal Interaction System for Children*

LI Jie, TIAN Feng, WANG Wei-xin, DAI Guo-zhong

(Laboratory of Human-Computer Interaction and Intelligent Information Processing, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

E-mail: lijie@iel.iscas.ac.cn

<http://iel.iscas.ac.cn>

Abstract: A pen/speech multimodal interaction system for children is designed and implemented in this paper. The system contains an integration framework used to integrate the information input by children from pen and speech. Children can communicate with the computer in a natural way via the support of several pen/speech interaction techniques, which are defined by the system. Children can draw 3D scenes and creative animals through the use of pen and communicate with the scene and the entity via pen and speech detection.

Key words: multimodal; children; gesture; speech detection; interaction primitive; interaction task

* Received March 5, 2002; accepted June 5, 2002

Supported by the National High-Tech Research and Development Plan of China under Grant No.2001AA114170

第 9 届 Rough 集、模糊集、数据挖掘与粒度计算国际学术会议(RSFDGrC 2003)

征文通知

RSFDGrC 2003 是 RSFDGrC 系列国际会议的第九届会议, 该系列会议每两年举行一次。RSFDGrC 2003 是该系列会议第一次在中国举办。RSFDGrC 2003 将由国家自然科学基金会、重庆邮电学院等有关单位共同主办, 于 2003 年 5 月 26 日~29 日在重庆召开。会议将针对 Rough 集、模糊集、数据挖掘、粒度计算等计算智能的理论与应用进行学术研究与讨论。

一、 征文范围

Rough 集理论及应用	计算智能	机器学习	文字计算
Fuzzy 集理论及应用	粒度计算	软计算及其应用	进化计算
Petri 网	软计算的逻辑基础	非经典逻辑	神经网络
软计算复杂性	空间推理	统计推理	智能 Agent
多标准决策分析	决策支持系统	知识发现与数据挖掘	多 Agent 技术
网络智能	集成智能系统	近似推理与不确定推理	数据仓库
模式识别与图像处理	其他有关领域		

二、 征文要求

- (1) 论文未被其他会议、期刊录用或发表;
- (2) 大会工作语言为英语;
- (3) 来稿一式 4 份;
- (4) 为联系方便, 请务必提供作者的姓名、单位、通信地址、电话、传真及 EMAIL 地址;
- (5) 论文版面格式及投稿信息请参看会议主页。

三、 重要日期

特邀小组会议申请: 2002 年 11 月 10 日(收到日期) 征文截止日期: 2002 年 12 月 10 日(收到日期)
 录用通知日期: 2003 年 1 月 10 日(发出日期) 提交正式论文截止日期: 2003 年 2 月 10 日(收到日期)

四、 联系方式

联系地址: 400065 重庆市南岸区黄桷垭镇堡上园 1 号 重庆邮电学院计算机科学与技术研究所

联系人: RSFDGrC 2003 程序委员会 王国胤教授

联系电话: 023-62460066

传真: 023-62461882

E-mail: wanggy@cqupt.edu.cn

[Http://www.cqupt.edu.cn/rsfdgrc](http://www.cqupt.edu.cn/rsfdgrc)