摘 要

多通道交互技术利用人的多个感知通道和控制行为的并行性，扩展了输入输出的带宽，提高了交互的自然性和灵活性。本文对多通道交互技术在教学中的应用进行了研究和探索，提出了在几何教学中将笔输入、语音与鼠标键盘相结合的构想，并最终通过一个原型系统——面向中小学的几何学习系统的开发，深入研究了应用这些技术的若干问题和实现方法。

当前教师在利用电子白板等手段进行电子化教学时，大多使用的还是传统的WIMP（Window, Icon, Menu, Point Device）界面。本文在多通道交互相关理论的指导下，以手写屏、麦克风、电子白板等工具，开发了更适合多通道交互的软件系统。该系统按照PIBG范式（Physical, Icon, Button, Gesture）设计，利用中科院笔输入平台和微软语音软件开发包开发，将语音与笔输入结合，使之成为笔交互的有效辅助手段。在系统设计中，我们将以用户为中心的场景设计方法，引入到多通道人机界面的设计当中，为可用性软件的开发做了一定的探索。此外，本文对信息的融合策略从任务结构描述、并行处理方面做了研究。本文的另一项主要工作是将几何图形识别完全融合到笔输入系统当中，使汉字识别、图形手势和命令手势识别结合。几何识别过程中几何特征与笔画数目、顺序无关。

本文受到国家863高技术项目（2006AA01Z328）和中科院计算机科学国家重点实验室开放基金（SYSKF0704）资助。

关键词： 多通道 笔交互 语音识别 多笔划图形

The Research and Implementation of the Geometric Learning System based on Multimodal Interaction technology

ABSTRACT

Multimodal interaction technology makes full use of the parallelism of the various perception and control action, expands input and output bandwidth, and improves the naturality and flexibility of interaction. This paper researches and explores the application of multimodal interaction technology in teaching, and combines pen input, speech recognition with mouse and keyboard. Ultimately, through the development of the geometric leaning system for primary and high school, this paper deeply researches the problem that how to apply these technology.

During the research, the author found that the whiteboard most used the WIMP(Window, Icon, Menu, Point Device)interface when teachers used computer in the classroom. Under the guidance of relative theory of multimodal interaction technology, this paper makes use of handwritten screen, microphone, Whiteboard, and other tools to develop a software system which more Suitable for multimodal interaction. This system, which Combining voice with pen input, is designed According to PIBG paradigm(Physical, Icon, Button, Gesture), developed geometric learning system based on pen input platform of Chinese Academy of Sciences and Microsoft Speech SDK. It makes speech recognition to be an effective accessorial interaction means of pen input. We Import the scenes design method to the process of multimodal human computer interface design, and do some exploration of availability software development. Moreover, the structure description of the task and the parallel algorithms are studied for the strategy of the information integration in this paper. Another important work in this paper is to integrate geometric shapes recognition into pen input system. This recognition method combines geometric shapes and command gesture recognition with characters recognition. The course of recognition is independent of the amount and order of strokes.

This paper was supported by the National High-Tech Research and Development Program of China (863 Program) (No.2006AA01Z328) and the Open Foundation of

State Key Laboratory of Computer Science, The Chinese Academy of Sciences (No. SYSKF0704).

Keywords: Multimodal, Pen-based Interaction, Speech Recognition, Multi-stroke Shapes

西北大学学位论文知识产权声明书

本人完全了解西北大学关于收集、保存、使用学位论文的规定。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版。本人允许论文被查阅和借阅。本人授权西北大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时授权中国科学技术信息研究所等机构将本学位论文收录到《中国学位论文全文数据库》或其它相关数据库。

保密论文待解密后适用本声明。

学位论文作者签名：王爽 指导教师签名：李安一

2008年6月18日 2008年6月20日

西北大学学位论文独创性声明

本人声明：所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含为获得西北大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：王爽

2008年6月18日

第一章 绪 论

1.1 研究的内容及意义

人机交互(Human-Computer Interaction)是研究人、计算机以及它们相互影响的技术。当前，得益于语音识别、自然语言理解、手势识别、计算机视觉等多门相关技术的发展，多通道人机交互(Multimodal Human-Computer Interaction)被越来越广泛的应用于各个领域。

目前，大多数教学软件使用的还是传统的 WIMP 界面（Window，Icon, Menu, Point Device）。WIMP 界面是目前主要的人机交互方式，但是随着信息多样化和信息量的剧增，这种界面范式的缺点也日益显示出来[1]：用户进行人机交互的手段是鼠标和键盘，如果用户界面需要人去适应计算机，就会增加人的认知负荷；不可避免的产生了低速信息输入与高速信息处理之间的矛盾[2]；在某些特殊的场合，如几何画板或某些制图软件，鼠标并不能直接高效的作图。这种键盘鼠标的交互方式，虽然在牺牲效率的前提下能做出较为精确的图形，但在某些并不需要精确制图的场合就显得极其笨拙。而以虚拟现实为代表的计算机系统拟人化和以掌上电脑为代表的计算机微型化和随身化，将是计算机的发展趋势。多通道交互技术就是在这种背景下发展起来的。它基于手写输入、语音输入、视线跟踪等多种交互技术，通过用户自身的感觉和认知，以并行的、非精确的方式与计算机交互。

发展至今，如何使计算机更加人性化，使计算机去适应人，而不是人去适应计算机是 HCI 今后面临的主要任务。多通道人机交互中，由于手写屏、语音设备等交互设备并没有按照一种协同工作的方式进行设计，更没有相应的应用程序以一种统一的方式，把信息流整理并告诉计算机，所以用户并没有真正的体会到多通道交互方式的便利。除此之外，在多通道交互中，笔输入和语音输入在交互上有不确定性。多通道研究的主要问题是对各个通道精确和非精确的信息进行整合，捕捉用户的交互意图，提高人机交互的自然性和高效性，最终使交互方式满足以用户为中心的要求。

1.2 国内外的现状

当前的人机交互作为计算机系统的一个重要组成部分，是计算机科学、心理学、认知科学和人素学（Human Factors）的交叉研究领域[3]，也是计算机行业竞争的焦点从硬件转移到软件后研究的新领域。

近 20 年，多通道作为人机交互研究的新领域在欧美越来越受到重视。在美国国家关键技术研究计划中，人机界面被列为 6 项关键信息技术之一。麻省理工学院 SLS（Spoken Language Systems）研究小组 GALAXY 项目为在线信息提供语音界面，已经应用于航班信息，天气预报，城市地图等查询服务。卡耐基-梅龙大学 ISL 实验室的 INTERACT 项目，期望通过多个通道（脸表情，唇读，手势，语音，视线跟踪）的处理和结合来增强人机的信息通讯。欧共体的 ESPRIT 计划也设立了 Amodeus-2 和 MIAMI 等多通道研究项目，主要研究用户与系统交互的模型、结构、表示和整合。比较著名的系统还有 Dynomite[4]系统，该系统使用笔和语音双通道来记笔记。此外美国从 70 年代就开始语音识别的研究，经过近 30 年的探索，语音识别技术经历了从最初的特定人、小词汇量、非连续、非独立扬声器到非特定人、大词汇量、连续、独立扬声器的发展历程，而且识别速度和准确率有极大提高。

我国多通道研究起步较晚，主要在语音识别和手写识别方面做了不少工作，近几年在一些科研项目如自然科学基金、863 计划、“九五”计划等的支持下进行了相关课题的研究。中科院软件研究所提出一种基于手势和语音的界面体系结构，提高了草图绘制建模的自动化与可重用性。我国在单通道界面研究方面同样也做了不少工作，如中科院人机交互技术和智能信息实验室，在笔式用户界面方面取得了很好的成绩，如实现了笔式用户界面平台（PIBG 工具箱）[5][6]。本文设计的系统中部分模块就是借助此平台开发。

1.3 问题的提出和本文的工作

传统的教学方式中，教师在使用粉笔时造成大量的粉尘，极大的危害了教师和学生的身体健康，更不利于环境保护。其次，由于大多数学习软件往往需要用

户在繁杂的菜单或按钮中寻找适合的命令，通过鼠标精确的定位来完成交互。这种输入设备和输出设备交互中非直接的操作，造成了某些信息输入的困难，比如图形的绘制就显得相当的笨拙。所以当前用户最需要的是好用、高效、具有充分表现力的软硬件系统来解决以上问题。

笔输入用户界面采用自然的交互方式，相对传统的 WIMP 界面具有非常明显的优势，如直接操作，简单灵活，而且笔输入的命令简明扼要，比描述性命令要好记忆，尤为重要的是它更加符合人们的使用和认知习惯。如果再加入语音交互进行辅助，来完成一些笔输入难以完成的操作，就可以更加准确高效的操作图形，并且语音的模糊属性可以有效降低用户的认知负担。多种交互相结合，也可以消除以往单通道交互的疲劳，使交互更形象更生动。

多通道技术在面向学习的交互系统中已经广泛应用。Takeo Igarashi 设计了一个基于勾画的 3D 绘画模型 Teddy~1。用户可以利用该模型自由随意进行 3D 建模。Teddy 主要采用笔和手势的交互方式，它不是一个精确的设计工具，而是生成粗略的 3D 模型。但它能够快速建模，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)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 [7]。这个系统采用 2D 投影的表现方法替代传统的 3D 表现方法，使绘制投影图景变得同传统 2D 绘图一样轻松。用户可以采用笔结合手势操作 3D 变换[8]。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软件 Doll Talk[9]。它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，对中小学生这一特殊用户群体而言，使用常规多通道交互系统又显得困难而且枯燥乏味[10]。其中的关键问题是小学生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。针对实体和场景的指定及描述过于精确对儿童用户是不必要的。

另一方面，由于计算机在学习中的广泛应用，教师可以利用电子白板等手段

进行电子化教学。但是传统菜单加按钮的系统界面显然不适用于电子白板，如用户很难用电子笔点击到投影在白板上方的下拉菜单。诸多问题限制了电子化教学的普及。此外，学生使用的学习软件，其系统界面大多都陈旧呆板，不能培养其学习的兴趣，也不能提高其学习的效率，与传统真实的书本学习并没有太大的差别。针对以上问题，笔者利用多通道交互技术，开发了面向中小学的几何学习系统。

本课题受到国家 863 高技术项目（2006AA01Z328）和中科院计算机科学国家重点实验室开放基金（SYSKF0704）资助。其中本文工作的特点主要是将理论和实践相结合，在原型系统的开发中做了一定的理论和算法探索，其特点体现在以下几个方面：

1. 将几何图形识别完全融合到笔输入系统当中，使汉字识别，图形和命令手势识别结合，做到几何识别过程中几何特征与笔画数目、顺序无关。

2. 将语音与笔输入结合，并运用一定的融合策略处理多通道的交互信息。

3. 将以用户为中心的交互场景设计方法引入到多通道人机界面的设计当中，目的是为可用性软件开发做一定的探索。

4. 最后，笔者提出以笔、语音以及鼠标键盘为交互手段，以电子白板和手写屏为大尺寸显示面板，开发了面向中小学的几何学习系统。这也是将多通道交互技术用于教学的一次探索。

1.4 本文的结构

本文首先论述了人机交互的相关理论，然后过渡到多通道交互技术的相关知识，最后，在 PIBG 工具箱以及语音识别库的基础上实现了一个原型系统，并提出了现有工作的不足和未来的改进方向。余下各章由四部分构成：

第一部分：由第二章和第三章构成，介绍系统开发所涉及的相关知识。第二章，人机交互技术研究。首先介绍了交互设计的原则和目标，以及心理学方面的相关知识。然后详细介绍了笔交互技术和语音识别技术。第三章，多通道交互的相关技术，主要介绍原型系统中涉及的关键技术。首先介绍了国内外对多通道技术的研究，随后介绍了多通道交互的优点，最后介绍了结合语音的图形用户界面。

第二部分：由第四章和第五章构成，介绍系统开发中两项独立的工作。第四章，基于笔交互的手势识别算法研究。对手势识别算法的实现是本文的主要工作之一。本章首先对整个笔交互框架作了简要介绍，然后重点介绍了系统对几何图形手势和命令手势识别的实现。第五章，多通道交互信息的融合策略。本文的另一项主要工作就是语音和笔输入信息的融合。首先介绍多通道信息融合的概念，然后描述了交互原语的设计，最后用实例的方式详细介绍了融合的整个过程。

第三部分：包括第六章，几何学习系统的设计和实现。主要介绍了论文期间原型系统的开发工作。不同于传统的软件开发，本章介绍了在需求获取时以用户为中心的场景设计方法。然后详细介绍了基于 PIBG 交互范式的系统界面开发。接下来描述了该系统的总体结构。随后介绍了使用的语音识别库以及语音识别流程。最后简要的介绍了系统的使用。

第四部分：第七章，总结和展望。主要讲述了研究的意义和工作的不足，以及对以后工作的展望。

第二章 人机交互技术研究

交互设计作为一门关注交互体验的科学产生于二十世纪八十年代，它由IDEO的创始人之一比尔.莫格里奇在1984年第一次设计会议上提出。从用户角度来说交互设计是一种如何让产品易用有效而且让人愉悦的技术。它致力于了解目标用户和他们的期望，了解用户在与产品交互时彼此的行为，了解人本身的心理行为特点，同时还包括了解各种有效的交互方式，并将它们进行增强和扩充。

2.1 交互设计的原则

交互设计涉及到了很多其他学科，包括：认知心理学、人类工程学、信息学、人因工程学、工程学、社会学、人类要素（HF）、认知学、认知功效学等等。设计者设计交互系统的交互机制与交互行为，目的是增强用户对该系统的体验[11]。优化人与系统之间的交互是交互设计的主要目标，这需要设计人员在设计系统时尽可能的支持用户的要求，满足用户的期望并且扩大用户的潜在需求，Norman总结了一些在产品的交互设计当中需遵循的基本原则：

1. 可视性。现代的交互系统中，功能的可视性越好，用户也就越容易理解交互进程。不可见的功能通常会对用户使用造成困难。

2. 反馈。反馈是可视性的相关概念，指返回的与活动相关的信息（包括已执行的动作或已完成的任务），以便用户能继续这个活动。反馈必须及时，若延迟超过了用户的忍耐限度，任务便很难进行。

3. 限制。表示在特定时刻用户的交互类型。通过对用户采用限制，可以主动防止用户误操作，客观上起了引导用户的作用，降低了错误率。

4. 映射。表示交互控制及其效果之间的对应关系。几乎所有的交互系统都存在这种映射，例如计算机键盘中的上下箭头分别表示光标的上下移动，笔式界面中的文本框代表一个文字输入区域。

5. 一致性。一致性指的是在设计界面时使用相似的操作，并且为相似的任务使用相似的元素。一致化带来的是遵循规则的界面，使其易学易用。

6. 启示性。启示性指的是事物的属性，即能帮助人们理解应如何使用这个

事物。例如：笔的“书写”动作是受笔的固有属性启发。如果一个对象的启示性是显而易见的，那么人们就很容易知道如何与它交互。启示性广泛应用于描述如何设计界面的对象，从而使得用户对应该采取的行动一目了然[12]。

2.2 交互设计的目标

1. 可用性目标

可用性目标可分为三类具体目标：易学性、灵活性和健壮性。

(1) 易学性

易学性是指新用户学习使用系统的难易。它包括系统的可预见性，即用户通过过去交互操作的经验来判断未来操作的效果；综合性，即用户按照系统的当前状态评估以前操作结果的能力；熟悉度，即新系统的知识范围应尽量贴近用户在真实世界或其他计算机环境中拥有的知识和体验；普遍性，即支持用户将他们对于特定操作的知识延伸到其他类似的情景中；一致性，即在相似的情景或任务中交互操作始终具有相似性。

（2）灵活性

灵活性是指用户和系统之间信息交流方式的多样话。它包括系统的对话主动性，即允许用户可以自由的摆脱系统对话形式上的限制；多线索对话，即系统支持用户同时进行多个任务的交互操作；任务可迁移性,即任务执行的控制权可以在系统与人之间相互传递，即可由一方主导又可由双方共同协作；可替换性，即要求相等的输入输出值可以相互替换；可定制性，即用户或系统可以根据具体用户特点修改界面形式。

（3）健壮性

健壮性是指对于用户成功地完成和评估目标的支持程度。它包括系统的可观察性，即允许用户评估系统的内部状态；可恢复性，即识别出过去交互的某个错误后达到目标的能力；响应度，即测量系统与用户之间的通信速率；任务执行，即系统在多大程度上以用户理解的方式支持用户要执行的任务，它包括任务对于用户意图的覆盖程度和任务被用户理解的程度[13]。

2. 用户体验目标

新技术已经从各个方面渗透到人们的日常生活中，在各应用领域，人们开始

对产品有了更多的要求。交互设计不只是提高工作效率，人们也越来越关心系统是否具备其他一些品质，仅仅用可用性目标不足以描述用户对交互行为的全部体验。

所谓“用户体验”指的是用户在系统交互时的感觉如何。用户体验目标不同于可用性目标，它更关注用户的主观感受，因此通常用主观性词语描述。用户体验的目标广泛应用与娱乐、游戏和电子竞技等行业，也是因为在这些领域中，产品的重要目标就是给用户带来心理上的愉悦。可用性目标更为客观，用户体验目标则更关心的是用户从自己的角度如何体验交互式产品，而不是从产品的角度来评价系统多有用或多有效。

2.3 认知心理学

如果要达到上述所说的交互目标，设计者必须考虑交互设计中的两个重要因素：信息的呈现和交互方式。我们可以从认知心理学的角度，对交互中的这两个因素进行评判。

从认知心理学的角度来看，人的认知处理能力主要受制于两个主要的因素：在处理过程中可得到的资源，以及可得到的数据质量。在针对某一个任务的认知处理过程中，充足的资源只是提高人的认知处理能力的必要条件，而不是充分条件。在可得到资源有限的情况下，资源数量的提高可以促进认知处理能力的提高。而当资源充足后，人的认知处理能力就只受制于可得到的数据的质量。

因此，在交互式设计中需要在资源和数据的质量之间找一个平衡点。由于大量资源的引入会给用户带来大的认知负担，从而增加用户的学习时间，增加用户的疲劳度和压力感，增加交互过程中的出错概率。Norman 曾提出通过提高数据的质量来减少资源的消耗。但数据质量的提高又依赖于用户对系统的训练和熟悉，这无疑要让用户化费大量的时间。如何在界面设计中解决这一两难选择是非常重要的问题。中科院软件所提出的PIBG交互范式主要就是为了解决以上问题。

2.4 笔交互技术介绍

在介绍多通道交互技术之前，我们先介绍手写输入和语音交互这两种单通道

技术。

2.4.1 笔交互技术的现状

从 60 年代初 Sketchpad[Sutherland 1963]作为第一个笔式用户界面系统问世，到目前各式笔式交互设备正逐渐步入人们的日常生活，包括 PDA、智能手机、电子笔记本以及功能日趋强大的 Tablet PC 等。基于桌面的 PC 是目前主流的计算设备，因此也是笔交互的主要环境之一。随着笔式用户界面的发展，手写设备也逐渐呈现多样性，目前以手写板、手写屏、声纳笔三种为主，如图 2.1。

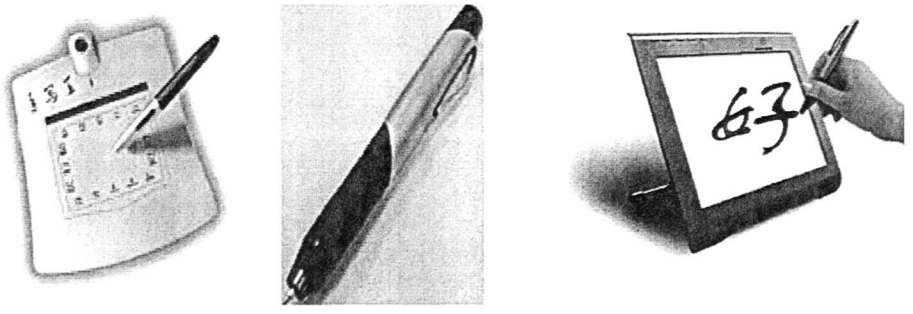


图 2.1 手写板 声纳笔 手写屏

此外在教学或者会议中也经常见到电子白板这种交互工具。它简单直观、也很容易的被人们理解和接受，在许多信息捕捉或信息交流的场合都得到应用。电子白板中提供给人们一种自由的、轻量级的大视角交互，同时创造了一种多人协作的信息环境，方便了人与人之间的交流。图中给出了 SMART 公司的电子白板。

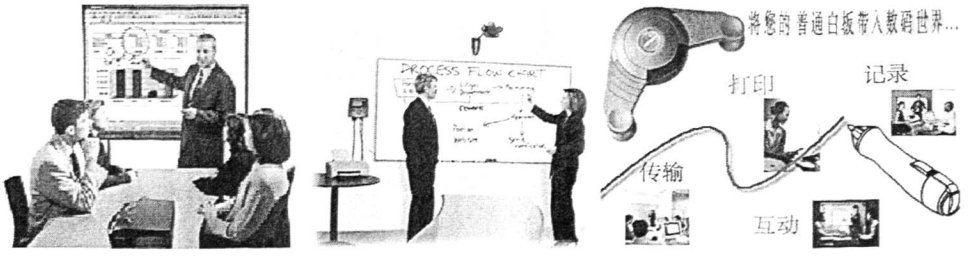


图 2.2 SMART 公司的电子白板

国外对笔交互技术的研究起步较早。华盛顿大学成立了专门的笔式计算实验室，研究内容包括笔式输入装置的结构和分类，笔式输入操作的评估，以及笔式用户界面。约翰霍普金斯的应用研究实验室正在开发一个基于智能笔的病历管理系统，该系统目标是让医生和护士都能通过一个笔式输入的掌上电脑输入和查询病人的病理情况，以适应医生移动办公的需要。

加拿大多伦多大学的“Haptic Research Group”把笔式输入作为重要研究方向之一。从理论到实践取得多项成果。如他们开发了新型 CAD 绘图系统。这种系统采用两手操作：左手拿十字光标器，操作显示器上透明且可移动的菜单，右手拿笔选择菜单或绘图。他们的研究一改传统的单手握笔或鼠标输入的模式，且暗示两只手操作笔会更加自然。

在日本，许多大小公司如 Wacom，Toshiba，Hitachi， NEC，SONY 等纷纷投资研究开发笔式输入技术。“笔式输入技术研究会”由日本东京电机大学发起，与 1993 年 7 月成立。会员们来自几所大学及十几所大公司的专门从事笔输入的专家，他们定期专门探讨笔输入技术，对产、学结合起了重要作用。东京电机大学人机交互实验室近几年一直注重 PDA 用户界面设计的研究，受到国际同行们的关注。

2.4.2 笔交互技术的特点

目前手写识别有两种，一是静态手写识别，就是把已经写好的文字以图像的形式输入计算机，让计算机对图像进行处理最后识别出文字；二是联机手写识别，这就需要书写的设备是实时连在计算机上的，输入计算机的数据是一个连续的与时间相关的坐标序列。两者相比，后者的数据中多了时间信息在里面。下面是对这种识别方式特点概述：

1.自然性

键盘和鼠标不是人类的自然交互方式，纸笔作为一个持续了上千年的传统工作方式，必然使人们感觉亲切自然。而笔式用户界面正是利用了基于笔纸（Pen-paper）隐喻的交互方式。

2.交互信息的连续性

在传统的用户界面中，用户发送离散的命令（如鼠标的点击）给系统，系统接收到命令后执行相应的任务。但在笔式用户界面下，信息的连续输入和连续反馈（笔迹）成为一个重要特征。在笔式交互中，笔划信息是用户输入的主要信息，它可以看成是一个连续的交互信息。它是一个时间段内在笔输入平面一系列点信息的连续输入。

3.非精确性

笔式用户界面的一个重要特征就是非精确交互，用户往往通过随意的勾画来完成交互任务。从交互设备上讲也是非精确性的，不像使用鼠标和键盘的 WIMP 界面，笔式用户界面下笔交互往往具有二义性。所以笔式用户界面中用户意图的提取和表示通常不是一个离散量，而是一个范围，或一个带有概率值的变量。

4.以用户为中心

在传统的人机系统中，人是操作者，人去适应机器。在笔式用户界面下。人们更加自由，因为笔式用户界面更加符合人们的习惯，如手势的利用。

5.交互的隐含性

所谓交互的隐含性是指用户在交互过程中不需要关注任务的执行方式和过程，而只需要考虑任务本身。这也是无处不在的计算追求的目标。从认知心理学的角度来看，这种隐式的交互方式具有非常大的优越性。笔式用户界面通过利用用户原有的（自然的）知识和技能，来提高用户在交互过程中的质量，由此来提高用户操作效率。同时将用户原有的知识和技能应用到交互中，无需或需要很少的训练时间，就可以帮助用户掌握界面的交互动作和思想。

6.基于手势的交互风格

手势是纸笔交互隐喻下的自然命令方式。用户无需转变思维方式，操作命令和操作参数直接蕴含在笔划中，简洁直观。基于手势的交互具有非精确、多通道、连续等特点，能够实现人的认知空间和计算机计算空间之间的平滑过渡，从而有效的改善了人机交互的瓶颈现象。基于笔的手势交互有很大的现实意义，对它的分析研究有助于推动计算机便携化的快速发展。

7.无处不在的计算

无处不在的计算（Ubiquitous Computing），有时也称为泛化计算（Pervasive

Computing)，是由 Weiser 在 1991 年首先提出的[14]。

手写输入的好处是显而易见的，不须专门学习与训练、不必记忆编码规则、安装后即可手写输入汉字，是最简单方便的输入方式。符合中国人的书写习惯，可以一面思考、一面书写，不会打断思维的连续性，是最自然的输入方式[15]。有些手写识别设备 (如汉王笔等)除了手写输入汉字外，还具有签名、绘图、保留手迹、替代鼠标等功能，这样既能实现手写识别也保留了计算机以前的输入方式。

2.4.3 笔交互技术的应用场合

笔交互技术从应用的角度主要可以分为：创造性设计工作（如概念设计）、信息交流和捕捉（如电子白板）、思想捕捉（如电子笔记本）和基于 GUI 的笔交互（如对遗产软件的笔交互增强）。这些分类之间没有严格的界限，因为一个原型系统往往具有几个分类的特征。以下给出这 4 个应用领域的简要概述和相应的系统说明。

1.创造性设计工作

笔交互的非精确性易于表达图形文字的特性以及自然的交互方式，使得它适于早期的、概念阶段的创造性工作[16]。因为创造性工作中，人们多进行抽象的、连续的思维。对于问题有一个模糊的认识，但不需要关心问题的细节，笔式用户界面对于这些活动有着良好的映射。它集中了笔式用户界面研究中的一大批著名的系统，其中 Sketch IT[17]是由 CarnegieMellon 大学机械工程系的 Tom Stahovich 设计开发的支持机械概念设计的工具。机械设计师在设计之前通常会在纸上进行概念设计，人们通过在纸上画一个特殊的例子帮助抽象思维。该工具可以将机械设计的草图转化为精确的几何描述，同时向设计者提供多个基于此草图的设计方案。

2.信息交流和捕捉

人们早已熟悉纸笔的工作方式。人们通过纸笔捕捉和交流信息是一个非常自然的活动。基于白板的记录和交流是纸笔工作方式的一个延伸，它允许在同一时间有更多的人参与到思想交流的活动中来，自然高效地实现了信息的共享。

Tivoli[18][19][20]是由 XeroxPAR 研究中心开发的用于非正式会议的电子白板系统，它运行在 Xerox 的电子白板 Liveboard 之上。该系统是 90 年代初期笔式用户

界面研究的第一个代表性原型系统，它提出了笔交互的一些基本概念，如笔划和手势。

Tivoli 提供给用户的是一个能够完成基本勾画任务的白板，它并没有对文字进行识别，而是保留了手写信息的原有外观。

Tivoli 是一个划时代的笔交互系统，它的出现为后来的笔式用户界面研究提供了新的思路，它也真正确立了笔式用户界面作为与 GUI 完全不同的一种界面范式而出现。

由 Georgia 理工、东京大学和 Xerox PARC 联合研制的 Flatland[21][22]是继 Tivoli 之后又一著名的电子白板系统，但它的目的并不完全是面向会议用途，而是针对个人办公室。它也是一种增强型的白板界面，为人们提供一种连续的、长期的工作方式。

3.思想捕捉

思想的捕捉在人们日常生活中，主要表现为做笔记，或者使用录音的方式。纸笔是做笔记的主要工具。通过纸笔，人们可以使用文字、图形、表格、大纲等多种信息表现方式捕捉重要的事件、想法，或者进行计划和安排。但是传统的方式一个缺点就是当要寻找历史记录时，比较困难，这就需要一些帮助。计算机刚好能满足该功能需求。还有就是结构化手写文档，帮助人们编辑文档。

4.基于 GUI 的笔交互增强

笔式用户界面在上述几类活动中发挥了重要作用，它还有一类应用就是对目前的主流界面 GUI 进行增强，如现在许多手写输入和图形编辑软件，也属于笔式用户界面。它们主要还是为了配合现有的 GUI 中的交互方式以及对某些遗产软件进行笔交互的增强，如 Palm 公司在其 PDA 产品 Palm Pilot 上的操作系统 PalmOS、微软公司的 Pen Coputing，它们基本上都在一个 GUI 的环境中嵌入笔式交互，而笔式交互也主要采用正式的（Formal）用户界面风格，即在交互过程中，将笔交互时间实时的转化为格式化的信息，它们关注于文字识别和基于表格的交互环境。这一类笔交互应用的研究并没有完全摆脱 GUI 的束缚，笔式用户界面的许多早期研究多为此类，而且它也是目前笔式用户界面在产业界投入市场的主要形式。

2.5 语音识别技术

语音识别是人机语音通信的一个重要组成部分，计算机语音识别过程与人对语音识别处理过程基本上是一致的，它是一个较困难的研究课题，问题本身涉及声学、计算机科学等许多学科。国内外在这个领域做了大量的工作才使得识别技术由实验走向成熟[23]。

2.5.1 语音识别技术现状

当前的语音识别系统主要可分为连续语音识别系统和孤立词语音识别系统。

连续语音识别系统是指用户用连贯自然的说话方式进行语音输入，而不必采用特定的、机器学习过的词语和命令。连续语音系统现在已经在医疗、国防等特定领域应用，但是这种系统现在仍有很高的错误率，而且开发的费用也很高，不能广泛应用，现在商业应用上更有实效的技术还是孤立词识别。

孤立词语音识别系统分为训练和识别两个部分。在训练阶段，用户将每一个词说一遍，并将计算得到的每一个词所对应的特征矢量序列作为模板存入模板库中。在识别阶段，将输入语音的特征矢量序列依次与模板库中的每一模板进行相似度比较，将相似度最高者作为识别结果输出。典型的孤立词识别系统可以使用几十个到几百个命令，尽管它并不像连续语音识别系统自然和易于使用，但是分离的命令还是易于学习的，而且具有比较高的准确率。

除了上述两种系统外，还有一种技术值得我们重视，那就是关键词提取技术。这种技术采用的是孤立词识别，但是却可以提供类似于连续识别系统的效果，使得交互更自然。典型的孤立词识别系统要求用户必须孤立的说出命令，在命令前后要有停顿。在这项技术中，用户可以说出一个包含待执行命令的完整句子和短语，系统将只保留下希望接收的命令而将其余部分过滤除掉。这种技术也可以被看作一个语法分析器，它可以让用户感到交互的过程更自然更直观，而其技术的实现难度要比连续识别系统小很多[21]。

2.5.2 语音在交互过程中的特点

1．语音信息难以保存。语音信号发出来以后，就不能再得到了，也可以说

语音具有一次性。因此，用户需要马上记住这些信息，消耗了用户大量的短时记忆资源，增加了使用者的记忆负担[24][25][26]。

2．语音是一种难于回溯和编辑的信息，它还会干扰人的其他感知通道。但是语音被证明在信息的前向处理上很有用，比如在紧急环境下的报警，为盲人和行动不便者提供输入和输出的途径。

3．语音对环境的依赖比较少，更不需要任何辅助设备，可以在空间狭小，照明不佳或接触不到交互对象等不良条件下正常使用，因此适合于视觉通道受到阻碍的的场合。

4．语音的效率很高，交互的信息内容十分丰富，而且接近于人的思维。如果人使用键盘进行文本输入，通过手输出想到的词语时，还会对他的话语和措辞进一步的琢磨和修正。

5．人对声音信号比较敏感，生活中人们经常利用声音进行提示和报警，比如比赛中使用发令枪命令比赛开始。此外，语音在信息随机呈现并要求操作员立即采取行动的任务中也非常适合。

6．人在进行肢体动作的同时可以说话，但不能在思考的同时讲话，这和人的大脑分工有关。例如人可以在走路和开车的同时进行谈话。因此人们发现在操作电脑时，人们可以在敲键盘和移动鼠标的同时进行思考，但却很难在说的同时进行思考。这是因为手眼的合作是由大脑的不同组织 (部分)完成的，可以进行并行的处理。因此语音在界面中可以单独使用，也可以结合鼠标、笔等指点式的设备进行交互。

7．语音对于说话的人效率较高，说话的表达方式比写字或者打字速度快，但是对于听众来说，听别人说话却比自己阅读要慢的多。与图形化的用户界面相比，语音界面是串行的输出方式，速度较慢[27]。并且不同于 WIMP 界面，可以执行的操作都可以显示出来供用户选择，对一个语音界面，如果没有适当的提示，对该系统陌生的用户可能感觉无从下手，不知道该说什么好，而且如果让用户记住所有的语音命令会增加记忆的负担。所以，语音识别系统必须注重语音界面的设计，这种界面可以使用的场合一般是有系统提示引导的问答式交互，而且每一步可供选择的项目不是太多或者是用户所熟悉的某一个领域的应用[21]。

2.6 其他单通道技术

除了上述两种交互技术，还有其他一些交互技术：

1.视觉跟踪：现在用户所使用的所有人机交互技术很多与视觉有关。早前的眼动跟踪技术仅应用于心理学研究，后来逐渐被用于人机交互。目前这种技术还处于起步阶段。

视线识别技术主要是解决眼睛运动特征的检测问题，目前主要的检测方法有接触镜法、电磁线圈法、红外光电反射法、红外电视法等。虽然鼠标键盘已经普及，但是对于某些特殊人群，如某些四肢麻痹的人可能无法靠鼠标来完成最基本的任务。对于这部分人群，如果他们能用眼睛来代替手操作，以后再加上机电控制技术就能够完全增加其独立操作的能力。另外在军事应用上，可以在飞行员的头盔中加入眼动跟踪技术，通过飞行员的视线定位所要打击的目标，这样可以减轻飞行员的操作负荷。

实现跟踪的基本工作原理是利用图像处理技术，使用特殊的摄像机对眼睛锁定，通过从人的眼角摄入和瞳孔反射的红外线连续的记录视线的变化，从而实现视线跟踪。另外从视觉追踪其读取的数据经过进一步的处理，最后提取出眼睛定位的坐标，这是一个复杂的过程，目前应用该技术的系统有 Applied Science Laboratories 制造的 Model 3250R 视线跟踪器[28]。

2.手势识别：手势是一种自然而直观的人际交流模式。基于视觉的手势识别是实现新一代人机交互所不可缺少的一项关键技术。手势交互技术不同于用手操作的交互技术，如鼠标器、键盘等虽是用手来操作，但这类设备比较简单，向计算机输入的信息基本上与手势无关，而且用户敲击键盘时是否遵循标准的指法，是单手击键还是双手击键都与输入结果无关。目前能识别的典型交互设备就是数据手套，它能对较为复杂的手的动作进行检测，包括手的位置和方向、手指的弯曲程度等，并根据这些信息对手势进行分类。

从手势识别的角度，我们可以将手势定义为："手势是人手产生的各种姿势或动作，它包括静态手势（指姿态，单个手型）和 手势（指动作，由一系列姿态组成）" [29]。由于手势的多样性、多义性、以及文化差异，手势识别是一门复杂困难极具挑战的学科，当前的手势识别研究现状及其应用可以从手势建模、手势分析和手势识别等三个方面去了解。当用户采用数据手套、摄像机获取视频图像后，首先分辨出每个手势，再对每个手势进行分析，最后是识别出该手势的表

示的意义。目前较为实用的手势识别是基于数据手套的，因为数据手套不仅可以输入包括三维空间运动在内的较为全面的手势信息，而且比基于计算机视觉的手势在技术上要容易。

3.表情识别：由于人面部表情在人交流中起着很重要的作用，它与语言一起使用，甚至可以使同样的文字表达出不同的含义。

如果要让机器认出人的表情，必须解决这样几个问题：第一，面部表情的跟踪；第二，面部表情的编码；第三，面部表情的识别。其中，面部表情的识别是最为关键的一步，需要丰富的表情数据库把读到的数据与模板进行对比，并且从中选取最佳匹配结果。

4.自然语言处理技术：自然语言处理不是一般地研究自然语言，而在于研制能有效地实现自然语言通信的计算机系统，特别是其中的软件系统，因而它是计算机科学的一部分。

我们知道，人与周围环境进行交互时获取信息可以是多通道的。人可以通过同时听一个人说话的语气和看他的面部表情及手臂动作来判断他的情绪。为了更好地理解周围的环境，人可以同时使用视觉、听觉等等。而普适计算的目标之一也正是希望人和计算机可以自然的进行交互。所以，对各通道的输入信息进行分析和融合以做出判断，就成为人机交互的一个研究方向。

此外还有三维输入、全息图像等交互技术，这些新交互技术使得系统交互通道的选择越来越丰富。

2.7 本章小结

本章首先回顾了交互设计的原则和目标，以及心理学方面相关的知识。由于多通道交互的基础是人机交互的单通道技术，所以接下来详细介绍了笔交互技术的发展，特别是对笔式交互技术的特点做了详细的叙述。然后本章又详细介绍了语音交互技术。最后概括的介绍了当今其他一些先进的交互技术。

第三章 多通道交互的相关技术

随着计算机技术的快速发展，人们已不满足于软件功能性的需求，而是开始追求自然的用户界面。

3.1 多通道交互的相关研究

2002 年国际标准组织 W3C（the World Wide Web Consortium）成立的“多通道交互”活动小组 MMI（Multimodal Interaction Working Group）致力于开发支持普适计算设备多通道交互的通用协议标准。我们所介绍的语音、手写等输入方式都在该标准中有明确的规定。其中“多通道交互框架”（Multimodal Interaction Framework）是最基础的规范和说明，图 3.1 是 W3C 给出的“多通道交互”的总框架。

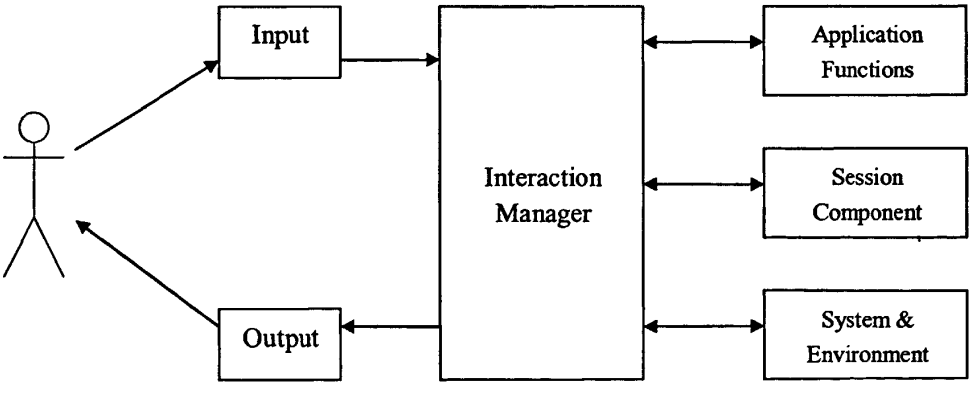


图 3.1 多通道交互框架

图中详细的描述了多通道交互的整体结构：输入构件、输出构件、交互管理构件、应用构件、会话构件和系统环境构件。而输入构件又分为识别构件、解释构件、集成构件等。

在美国，MIT 的多模式界面 AGENTS 项目包含了智能笔技术，该项目将智

能笔技术和其他的多模式输入技术如语音识别和表情识别等结合起来，以提高用户界面的交互效率和自然性。MIT 的 AI 实验室对于以勾画为特征的笔式用户界面进行了大量的研究，在机械设计和物理仿真等应用领域设计实现了笔交互系统，如 ASSIST[30]。

CMU 大学人机交互学院将笔交互嵌入到了工具箱系统 GARNET[31],同时设计实现了通过勾画设计图形用户界面原型的工具 SILK[32]。Berkeley 大学 GUIR 实验室开展了大量的笔式用户界面研究，其中以基于笔的设计工具为主，如网站设计 DENIM[33],同时也设计实现了支持比较户籍面开发的工具系统 SATIN[34]。

Georgia Tech 的未来计算环境实验室和无处不在计算实验室在以无处不在计算为背景，结合使用了大量上下文信息的基础上，进行了笔式用户界面研究，如 Flatland[35]。在 Brown 大学也展开了相应的研究，如基于笔的乐曲制作系统 Music Notepad[36]。

国内诸多研究机构也从笔交互的模型和体系结构、识别算法等不同角度对笔式用户界面进行了研究。如中国科学院软件研究所的人机交互与智能信息处理实验室在笔的字处理、概念设计和交互平台等方面进行了研究；北京大学在笔的多通道应用方面进行了探索；北师大心理系和中科院心理所在笔交互的认知机理和试验评估方面展开了工作；微软中国研究院在基于 Tablet PC 计算平台上设计开发基于笔的用户界面软件，同时进行了笔交互硬件研究。

3.2 多通道交互的优点

与单通道界面相比，多通道有以下一些优点：

多通道交互方式可以各取所长，发挥各个通道的优点，显著的提高系统的识别率。

多通道交互在功能上的交叉，在一定情况下可以互相取代，因此多个通道在系统中并存的冗余能够带来不小的鲁棒性，提供对环境与用户的适应性以及用户可选择性。用户可以选择自己所喜欢的或者更适合自己的通道。当一个通道由于故障或者环境的制约不方便使用时，用户可以选择其他的通道完成任务。多通道优势还在服务残疾人上。此时眼动界面对高位截瘫者的重要性，正如声音界面对

于盲人的重要性一样。

多通道交互能够提高输入/输出的带宽，而其实质是帮助人机顺畅交流。如果从外设与主机间接口上比特流的大小来看，每引入一种新的输入输出设备都会增加相应的比特流。引入语音将增加语音信号数据，引入唇读将要增加嘴唇的图像数据，这确实会提高带宽。

无处不在计算的提出启发我们可以将新的方法和技术应用于多通道交互设计中，它对不可见交互的强调以及经验捕捉都为新的交互方式研究提供了很好的借鉴和指导作用。

3.3 结合语音的图形用户界面

通过实践发现，单纯的笔交互界面还存在一些缺陷。在自然生活当中，人长时间使用纸和笔时最容易出现疲劳。而当前手写屏由于屏幕光滑，摩擦力小，长时间使用更容易造成用户的疲劳，加之手写笔操作上的复杂性以及舒适程度的欠缺，这种只靠笔交互的系统界面还不能满足用户自然、高效、舒适的要求。因此我们需要把传统的键盘鼠标交互方式，以及语音识别技术加入到界面设计当中来解决这个问题。

3.3.1 语音界面的特点

将语音和笔输入以及传统的键盘鼠标交互方式结合，有很大的优点。笔和语音的操作简单，易学易用，交互命令要比菜单按钮的命令方式容易记忆。对于图形操作，利用笔进行自由的勾画绘制，自然方便，而利用语音可以进行一些辅助性的交互动作，来完成笔很难完成的操作，消除笔输入在识别率上的歧义性。比如语音交互命令完全可以不受图形对象本身可见度的限制，从而可以更加准确的操作图形。另外单纯的笔交互缺乏生动性和形象性，尤其是面对中小学生这样特殊的用户，长时间单一的操作容易造成疲劳，注意力不集中等问题，笔输入和语音结合很好的解决了这个问题。

在这种多通道界面中语音界面不占主导地位，它只是界面的一部分，和其他形式的界面 (通常是图形界面)结合在一起。其实这是一个多通道界面，这种界

面可以完成任务的方式有以下几种：

1. 可以单独用语音完成。

2. 需要视觉或手动操作。

3. 至少需要语音和视觉通道合作完成。

典型的应用有：允许用户进行文本听写的字处理软件，允许用户用语音进行网络浏览的网络浏览软件。

3.3.2 语音界面的交互方式

人机交互是一个双向的过程，而且语音和其他的交互手段相比又有着许多独特的地方，所以在设计一个语音界面之前，我们应该首先建立一个模型，依据这个模型来设计和实现界面。语音界面的交互模型主要考虑以下几个方面。

1. 语音界面的心理模型：心理模型主要说明一个信息传输包括了消息、发送者、接受者和信道四部分。一个信息传输的过程包括概念/信息的产生、编码、传送、接收、解码和反馈六个阶段。

2. 语音界面的交互模型：信息交流的目的是使交流双方对于所交流的信息获得共同的理解，也就意味着发送者和接收者都要以同样的方式来理解信息[37]。图 3.2 所示是一个语音界面的交互模型。

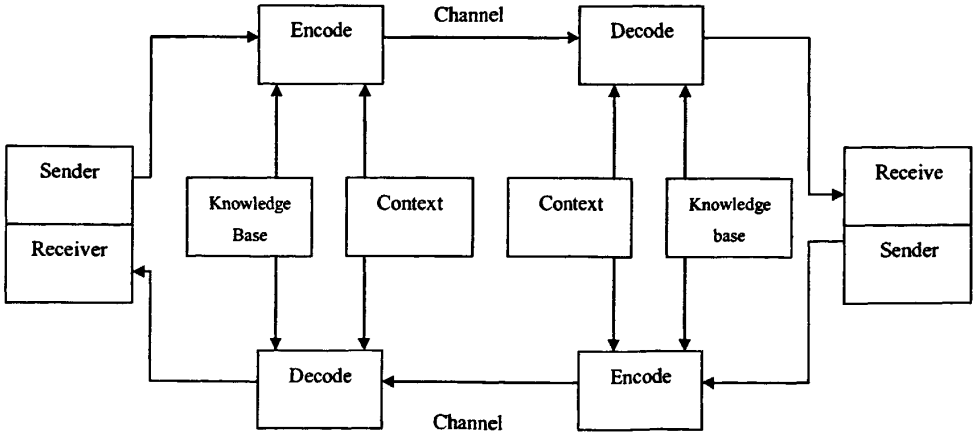


图 3.2 语音交互模型

3.3.3 语音界面中人的因素

在以往人机界面的设计中，设计者往往只考虑计算机系统的功能实现。其实界面的设计包含两方面内容：计算机技术和人的因素，只考虑计算机技术是以往软件系统设计的一个误区。计算机工作者研究的所有技术和产品，最终服务的对象都是人，因此在设计中一定要考虑到人的因素 。语音用户界面的设计也是如此。

人的因素就是指在进行产品设计时考虑到人的需求和特点。它的核心思想是在进行计算机界面的设计时将我们所知道的人的因素考虑进去，以提高系统的可操作性、操作速度，降低错误率和为用户提供更好的满意度[21]。现在，语音技术不仅考虑到人的生理方面的因素，还要考虑到人的认知和心理方面的因素。它有以下几个关键的问题：

1. 高错误率

系统产生错误时，观察用户的反映是很重要的。它能帮助我们寻找合适的补救方式。

2. 不可预知的错误

人的语言在人与人沟通时可以被正确理解，但计算机理解起来可能存在二义性。

3. 对于语音系统的期望

用户一直渴望与计算机交流能像与人交流一样简单，这也是设计者追求的目标。特别是那些不熟悉计算机，或者行动不方便的用户，他们希望计算机能像人一样听懂自己的命令，也能像人一样用语音反馈执行结果。

4. 工作于多通道下

很多时候，人是使用多个感知通道来完成任务的，如视觉、听觉和触觉等。

5. 单一的语音系统占用大量的记忆资源

由于单一的语音通道缺乏视觉反馈和确认信息，因此人们的记忆资源会被大量的占用。

6. 口语与书面语的差别

人与人交流时使用的是口语，因此语音系统要采用与书面语不同的命名规则。

7. 说话的方式可以被塑造

一个人可以很容易地影响另一个人的说话方式，语音技术就可以利用这个特性[37]。

3.4 本章小结

本章主要为下一章介绍基于多通道技术的原型系统做知识储备。首先对多通道技术的现状做了分析。然后详细列出了多种交互通道结合的优点。由于基于语音的人机界面的特殊性，本章随后介绍了结合语音的图形用户界面。

第四章 基于笔交互的手势识别算法研究

手势识别和文字识别是笔交互两个重要的组成部分。由于笔交互的不精确性，所绘制的图形大多是不规整的草图，需要系统识别并加以修正。而书写的文字也只有识别后，系统才能得到文字的内容并加以处理。

几何学习中，用户经常要进行几何图形的绘制。因此我们在开发原形系统时，对手势识别算法进行深入的研究，使系统能够快速准确的识别几何图形以及命令手势，并和文字识别结合在一起处理用户在几何学习中的交互。

4.1 笔交互任务生成框架

系统对语音输入和笔输入这两个通道输入信息的识别是独立进行的，因此在介绍系统对输入的处理时，需要把语音模块和笔输入模块独立介绍，这也是系统最为复杂的两个模块。

该系统的设计风格采用 PIBG 交互范式，笔交互任务的生成部分基于中科院的 PIBG 工具箱开发。如图 4.1，工具箱采用分层的方式来设计统一的任务生成框架。框架共有四个层次：硬件层、交互信息产生层、交互原语构造层和交互任务构造层[5]。

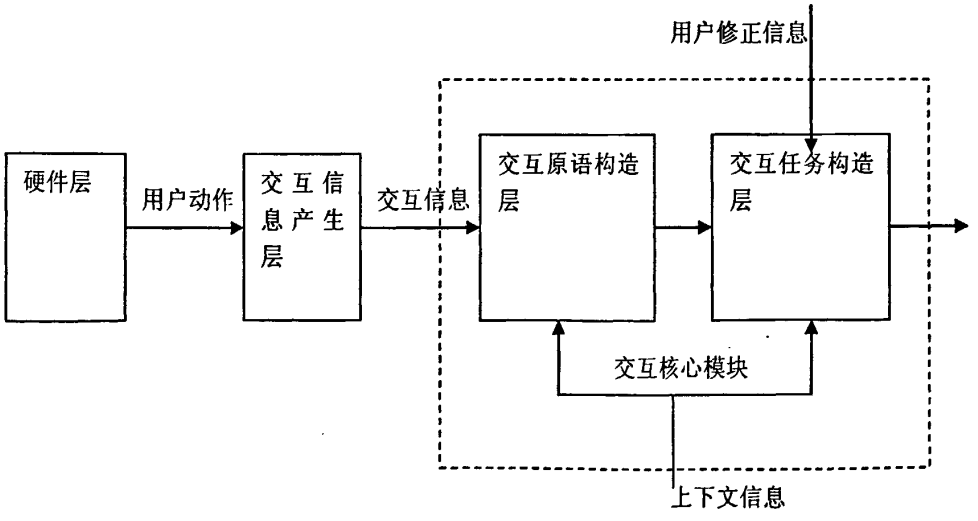


图 4.1 笔式交互任务的生成框架

其中我们又将交互的信息分成两个层次：消息层，原语层。

（1）消息层

消息层中的信息是指系统直接从硬件驱动中得到的消息。 目前共五种，分别指 windows 左右键发送的消息：

LB_D：左键按下

LB_U：左键抬起

B_M： 鼠标移动

RB_D：右键按下

RB_U：右键抬起

目前正在扩展第六种消息：PenEvent, 此消息中除了包含上述六种消息所包含的位置信息之外，还包含压力和方向等信息。

（2）原语层

原语层中的信息是指经过处理后，在笔交互中基本的交互原语。它是来自底层的独立的、最小的、不可分割的操作。

4.2 手势识别

手势是 PIBG 范式下用户和系统之间交互的主要方式（对于 PIBG 交互范式我们将在 6.2 节详细介绍），作者对工具箱中的手势做了一些补充和修改。

手势识别是手势设计的重要组成部分。笔交互界面对桌面系统和便携式系统都能够提供很好的支持。特别值得指出的是，用笔来完成的操作命令，即手势很受人们欢迎。实践证明无论是在很小的操作空间很有限的设备上，还是在很大的操作区域，甚至超过人们手臂控制范围的较大屏幕上，手势的应用都是很广泛的。

目前识别算法有很多，常用的有基于神经网络的识别算法和基于特征的识别算法。神经网络识别算法识别率较高，但这种识别率是建立在拥有大量训练样本的基础之上的，这种算法在实际应用中很少被重复设计。基于特征的识别算法比较简单，它的每个手势类只要求很少的训练样本。目前这种算法已经成功的应用在了很多系统上。系统中的手势集合由手势类组成，每一类是一单独类型的手势，

一个类由一系列训练样本来定义，而每个手势类至少需要一个样本。识别器正确地判断绘制的手势属于集合中的哪个类。基于特征的识别算法中一些相关特征包括初始角度、总长度等，据此计算出不同的匹配概率。但由于识别算法限制，很多系统目前只支持单笔划手势识别，这就要求设计的手势要尽量简单。但在很多应用中，交互中的操作比较复杂，需要设计较多的手势，所以多笔划手势的识别就显得相当重要。

4.2.1 手势的识别流程

笔划手势可以让用户通过一个笔划或者少量笔划明确地描述一个操作或表达特定信息，其中笔划指单个的落笔—>移笔—>提笔的过程（或者鼠标的按键—>移动—>松键过程）。支持单笔划识别的系统，用户必须在一笔之内将想要让系统识别的图形绘制完成。本系统采用多笔划识别，整个过程可以有多次落笔、抬笔，不仅更贴近自然的绘制方式，方便了用户，而且也大大的增加了手势识别的种类。

本系统将手势分为图形手势和命令手势两大类，系统识别计时器判断用户手势绘制完成后，将手势信息发送给手势分类器。分类器分离出几何图形手势和命令手势，由识别模块分别进行识别。

为了实现功能性需求和易用性需求，系统将主要的笔输入通道分为四大模块来实现：笔交互数据收集模块，原语构造模块，识别模块，语义构造模块，如图4.2。

在输入层：数据收集模块负责收集笔交互输入的数据及时间信息，每个笔划的起始点和终点以及用时最长的点。在识别层：原语构造模块对收集到的笔交互信息集合（从笔压下到笔抬起）及时间信息进行融合，产生相应的笔划原语；识别模块负责识别几何图形，命令手势以及文字。在应用层：语义构造模块基于识别后的几何图形及上下文信息生成特定领域中有意义的实体，及将识别出的命令映射为相应的操作。本系统采用 XML 语言来定义数据存储结构，识别后图形的所有数据以 XML 格式进行存储，为进行广泛的信息共享提供了基础[38]。

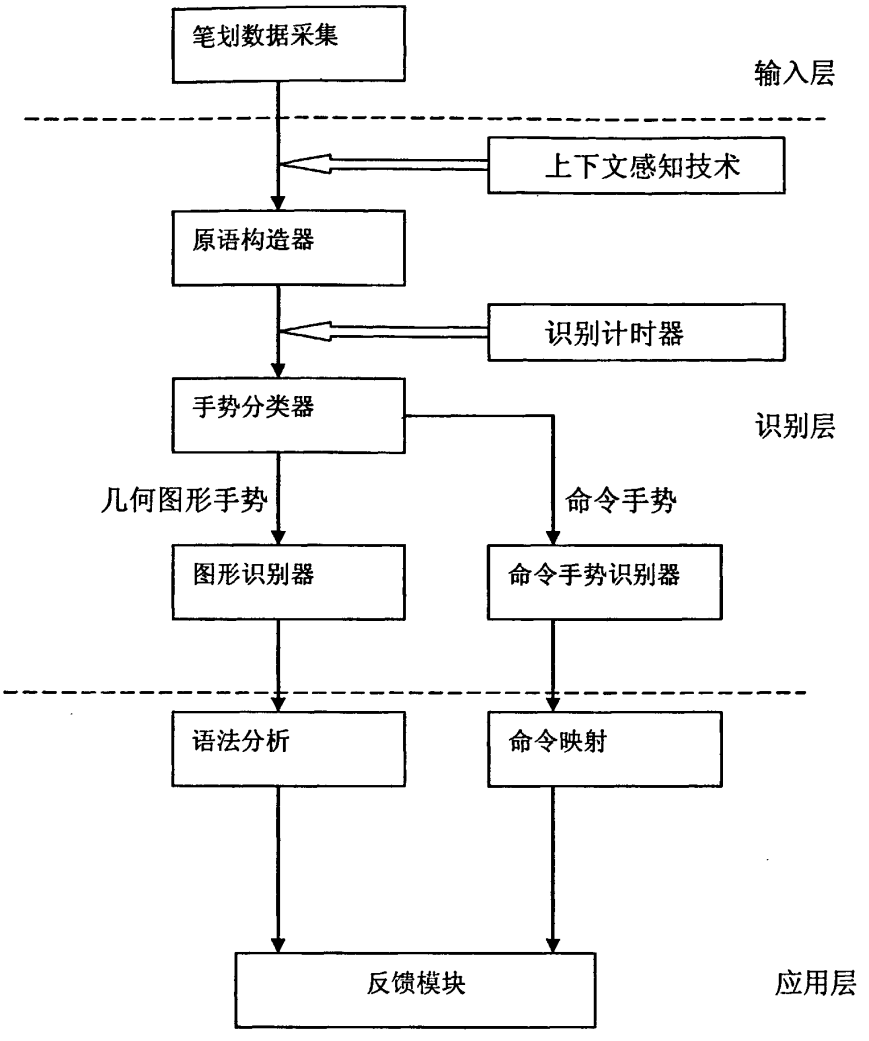


图 4.2 系统体系结构图

用户画笔划过程中，系统实时采集原始笔划信息，并提取每一笔划的起始点，终点以及用时最长的点，以用于手势识别的特征点。

用户画笔划过程中，每一笔原始笔划的信息送往原语构造器（PC）中，在手势识别状态中，进行笔划类型的判断：

if 笔划类型为顿笔笔划

if 第一笔划

then 置手势标志为命令手势，保存笔划信息；

else 清空所有笔划信息；

else if 笔划类型为普通笔划

if 第一笔划

then 置手势标志为几何图形手势，并保存笔划信息；

else 将笔划信息保存；

else 将所有笔划清空；

判断笔划类型为有意义笔划（即为普通笔划或顿笔笔划且为第一笔划）后，开启识别计时器开始计时。用户落笔前，超过设定的时间，就调用手势分类器开始识别。

4.2.2 图形手势识别

图形识别算法主要基于以下思想：

1．为了支持用户自由的作图，所选用的几何特征必须与笔划数目，顺序无关。

2．为了使手绘图形的识别与大小无关，采用全局特征的比值作为决策因子。

3．为了减少手绘过程中引入的噪音影响，所用到的几何特征均为全局几何特征。

4．为了克服几何图形手势的不确定性既非精确性，采用了模糊逻辑来增加识别图形的确信度，阀值的取值依赖于经验值的判断。根据此算法，系统能够很好的识别直线，矩形，圆，椭圆，三角形，如图 4.3。经过试验，该系统对直线，三角形和矩形的识别率均在 90%以上。对圆的识别率为 71%，对椭圆的识别率为 85%。

从第一笔按下事件到最后一笔抬起事件经过一定时间间隔，通过收集笔划信息开始识别过程。接下来，计算输入点集的凸包。利用凸包计算两个特殊的多边形。使用一个简单的三点算法识别出凸包的最大面积的内接三角形和外接矩形。最后，计算出每一个多边形的面积和周长以便评估其特征以及与每个图形的相似度。

为了从图形中辨别出直线和圆，采用决策因子 Pch2/Ach。这里的 Ach 是凸包的面积，Pch2 凸包周长的平方。圆的决策因子是最小的，因为它是包围一个给定区域的具有最小周长的平面图形，这个比值大致为 4π。另一种极端情况，

直线的决策因子接近无穷大。

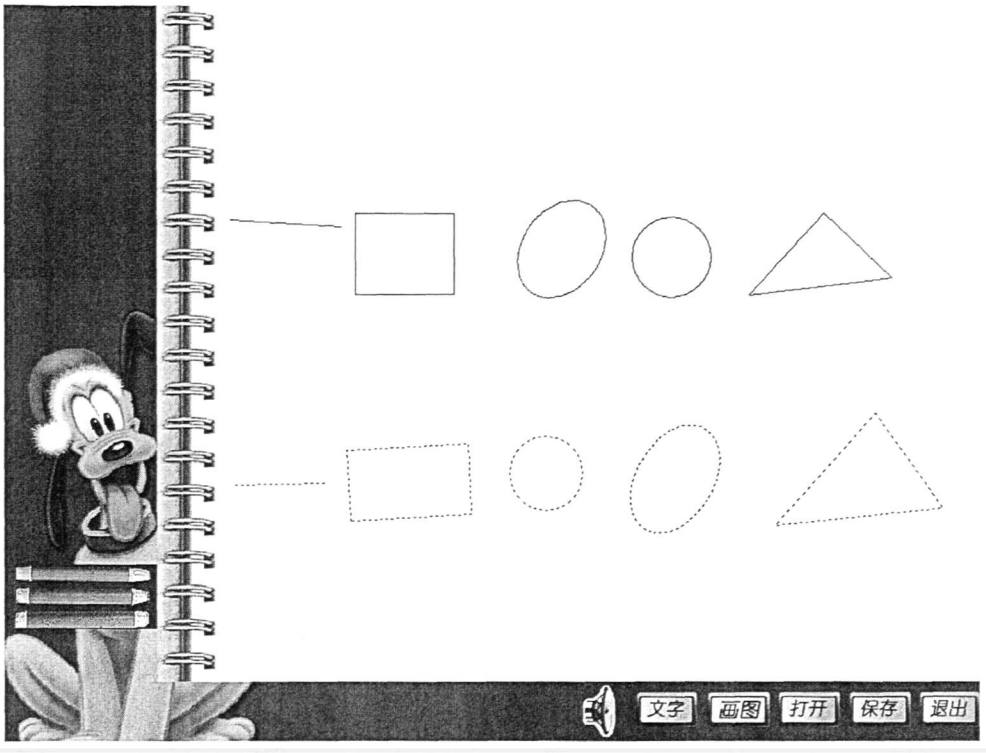


图 4.3 系统识别的直线，矩形，圆，椭圆，三角形

三角形的识别是通过凸包的最大内接三角形的面积（Alt）与凸包的面积（Ach）之比来判定。三角形（Alt/Ach）比值趋于 1，其他图形（Alt/Ach）比 1 要小。

类似地，Per 为外接矩形的周长。对于矩形，外接矩形的周长与凸包的周长特别接近，因此通过 Pch/Per 把矩形与椭圆区分出来。

因为椭圆最难辨别，因此用更复杂的比，其中包括凸包，最大内接三角形和外接矩形的面积。这个比（Ach2）/（Aer*Alt）若趋近于 1，且面积比（Alt/Aer）的值高于 50%，那么这个图形按椭圆处理，否则该图形不能被识别。

对于每类几何图形，都有一个决策因子，它的取值范围与其他图形偏离比较大，利用这个特点，使其与其它图形分离开，将此图形识别出来。对于每类图形所采用的决策因子的数目是不同的。而这些决策因子的作用也是不一样的，其中有一个用来识别几何图形，而其他的是避免错误的识别结果。

为了识别某一类几何图形，用到的每个决策因子都有一个针对该类图形的取

值范围，这些取值范围的集合我们称为模糊集。而模糊集来自于精心的实验及仔细分析的结果。

下边给出一个识别矩形的例子：

if Pch2/Ach IS NOT LIKE Line AND

Pch2/Ach IS NOT LIKE Circle AND

Alt/Ach IS NOT LIKE Triangle AND

Pch/Per IS LIKE Rectangle

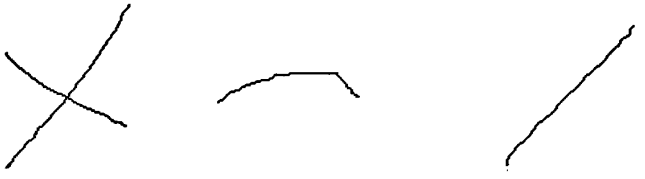
THEN

Shape IS Rectangle

由于该原型系统只是基于全局几何特征（凸包面积，凸包周长，外接矩形面积，外接矩形周长，最大内接三角形面积）进行几何图形的识别，是一种基于手绘后静态几何图形的识别方法，与手绘过程无关。所以该识别方法与笔划数和顺序无关。

4.2.3 命令手势识别

系统可以识别 4 种手势：删除，撤销，移动和文字切换，如图 4.4 所示。



1. 删除　　2.撤销　　3.移动　　4.文字切换

图 4.4 系统识别的命令手势

其中，删除和移动的交互任务需要先选定操作对象。

此外用户可以通过点击进行绘制图形和文字输入的切换。

在命令手势的识别中用到了交互任务的概念。交互任务使用三元组[39]描述：<交互对象, 任务类型, 参数列表>。

系统根据识别的手势(或指点操作) 确定任务类型和参数列表，生成当前对象交互任务的操作原语。如当前对象为一直线 line1，手势为删除 delete(其数据表示为 NULL)，则得到“直线 line1 删除”交互任务三元组：<line1,delete,NULL>。

4.3 文字识别

除了图形绘制的功能，系统还具有文字识别的功能，如图 4.5。该功能保持笔迹的原形，并能对笔迹进行编辑操作。用户可以选定文字并对文字删除。这种识别方法既保持了笔迹的原形，又使这些离散的笔迹变成了有意义的结构化单元（具有字结构，行结构等特征），达到了修改、维护、检索的方便性。对笔迹的结构化识别，是西北大学人机交互实验室基于中科院的文字识别库完成的，有关这一部分的研究读者可以参阅中科院的相关资料[40],在此不作论述。

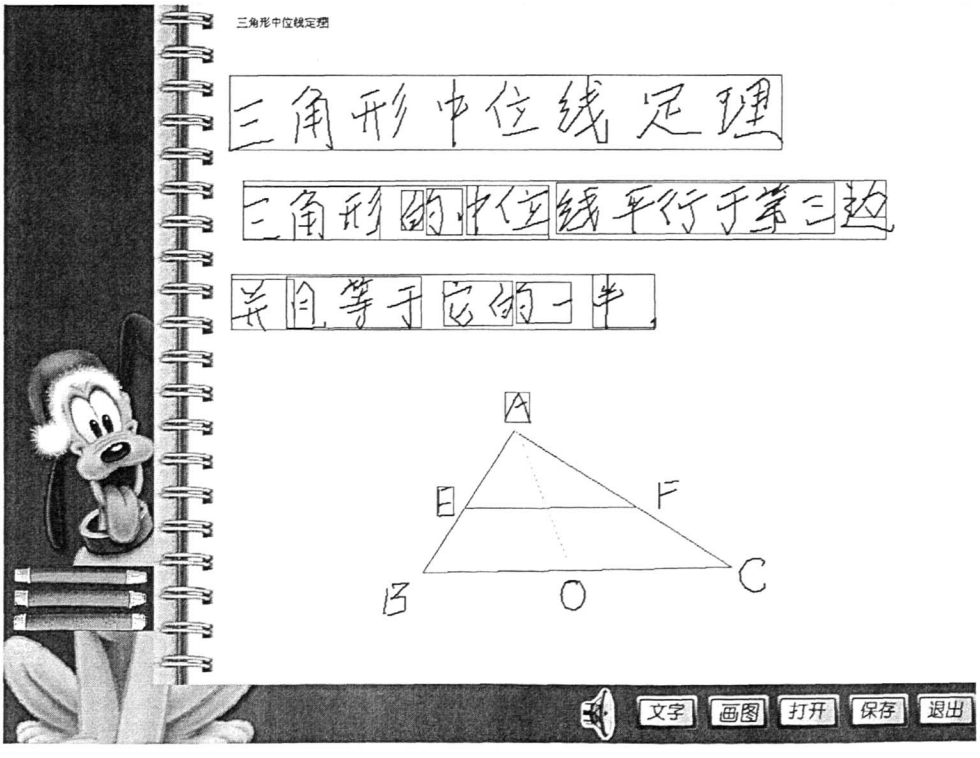


图 4.5 系统的文字识别

4.4 本章小结

图形和命令手势的识别是本系统的重要功能。本章首先概括的介绍了笔交互任务生成框架。然后重点描述了手势识别算法的实现。其中主要包括三方面内容：手势识别的意义及识别流程，图形手势识别以及命令手势识别。最后简单介绍了系统的文字识别功能。

第五章 多通道交互信息的融合策略

对交互信息的融合一直是多通道技术研究的重点，这也是本文的另一项重要工作。多通道交互的特点是描述性和直接性，与传统的鼠标键盘交互不同，这种交互方式并不追求过分的精确。如果交互信息足以完成一个交互任务，那么就不必对用户的输入多加限制，这样体现了多通道交互“充分性”特点。为了达到交互的“充分性”，多通道交互提倡多个通道协作工作，这样可以相互补充交互中单通道交互信息的不足[41]。同时多通道交互系统也要对这些并行的信息进行解释和整合，才能确定一个具体的任务。

5.1 多通道信息融合的概念

从人本身来说，感知信息的方式就是多样的。比如人和人交流的时候既使用语言，也使用手势、表情等方式。多通道人机交互的目的就是尽量让人按照习惯的方式来操纵计算机。其中的关键问题就是人的多通道怎么转换成计算机能够理解的形式，这一过程，我们称为多通道的融合[42]。

在处理融合问题时，人们提出了很多模型和算法。比如分层融合模型，它把融合过程分成词法、语法和语义层。在语法层，有多个通道传递意义相同的信息。在语法层，把词法层的原语信息按照人机交互的语法规范分成命令原语、对象原语、对象属性原语。在语义层，利用任务驱动机制，将原语组合成各种具体的任务。

此外，还有基于概率的指称融合模型，以及面向任务的多通道界面结构模型。这些融合方法的目的都是将原先不精确的输入信息通过整合，给出满足实际任务得以实现的充分性信息。以下是本文针对原形系统特点给出的融合方法[43]。

5.2 交互原语设计

系统针对笔和语音在交互中定义了几种交互原语，它是用户和系统在输入

输出上词法级的交互，是从各个通道得来的最小的、不可分割的操作。在交互过程中，底层的信息是交互设备输入的原始信息，需要在词法层对其进行统一化处理，把目的和意义相同但形式不同的信息，整理成为统一的结构进行表示。这些与底层无关的信息在一定的应用上下文中有着特定的交互意义。交互原语可以定义为[10]：

{交互动作，数据结构，使用通道，时间标签}

交互动作是用户与系统之间交互的具体行为，通常与交互任务相对应。数据结构表示描述行为的属性，使用通道指示交互行为所在的交互通道，用于多通道信息的融合。时间标签表明执行的时间戳，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。而不同通道交互行为时间上的接近性是建立跨通道关系的基本依据之一。

5.3 交互信息融合策略

用户通过语音、笔两个通道输入信息，系统将对来自两个通道的信息进行识别，识别后用交互原语的形式表达每一个通道提取出来的信息，这一过程分别在各个通道的识别模块内完成。信息被识别后，语义理解模块将两个通道的信息分别进行语义理解。来自不同设备的语义信息被分别处理，统一表示为各个通道的语义结构体，最终提取出与设备无关的信息。在多通道语义融合中，根据特定规则在上下文信息的支持下，融合来自语音和笔通道的语义信息。经过融合后的输出是包括明确、完整语义信息的三维任务，以及文本和语音的多通道输出。经过融合的三维任务既可以反映单个通道的语义信息，也可以反映多个通道的信息相互融合后形成的语义信息。也就是说 ，用户既可以通过多个通道的串行交替操作来进行交互，也可以根据两个通道时间的相关性进行并行的协同操作。在呈现给用户融合结果之后，用户还可以使用语音和笔对错误结果进行修改，最终形成正确的交互任务。

由于多通道交互的非精确性，我们对于语音交互中不能明确的信息，笔输入通道可以给出描述性定义。对于几何学习系统而言，用户一边绘制图形，一边可以用语音对图形操作。例如用户绘制一条直线，当直线被识别并显示后，

可以发出语音命令“移动”，这时需要笔通过点击来确定移动的目的位置。这个实体级的任务对应于两个交互原语 Penlocate 和 Spwords。

Penlocate:

<tap/gesture, locate, pen-based, time-stamp >

Spwords:

<speech command, words, speech, time-stamp >

Penlocate 以点击命令手势作为交互动作，以笔为使用通道，以中心坐标作为动作属性，同时生成时间标签；Spwords 以语音命令作为交互动作，以语音为使用通道，以语音内容为动作属性，同时生成时间标签。

首先系统识别出语音命令“移动”，根据约定等待用户接下来的手势命令。此时用户输入手势命令，系统生成时间标签。当两个交互原语生成，系统会依据时间标签将两个交互原语整合，构成一个完整的动作。这时操作对象，动作类型，以及参数都已确定，系统根据这些信息开始移动图形，直到图形移动到参数中指定的坐标，当前的任务完成。

5.4 交互信息的并行处理

系统采用分层的结构，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。系统主要整合语音和笔两个通道的输入信息，具体实现和处理流程结构如图 5.1 所示，归纳起来可分为 4 个步骤：(1)手写信号识别；(2)语音信号识别；(3)语音信号解释；(4)手势语言解释；(5)最终多通道统一执行解释。

由于交互信息来自于不同通道，且并行产生，系统采用了一种优先级的整合策略。这种策略基于对语音输入、笔输入信号以及指点设备信号的并行处理。在处理过程中，系统在识别手势信息的同时检测语音信息。系统结合应用运行时的上下文,针对输入信息分别并行给出手势语言解释以及语音信息解释，按照事先约定的优先级(优先级由开发者事先设定)，最终统一执行解释。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优先级整合策略，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扩展通道时均可保证系统

正常运行。

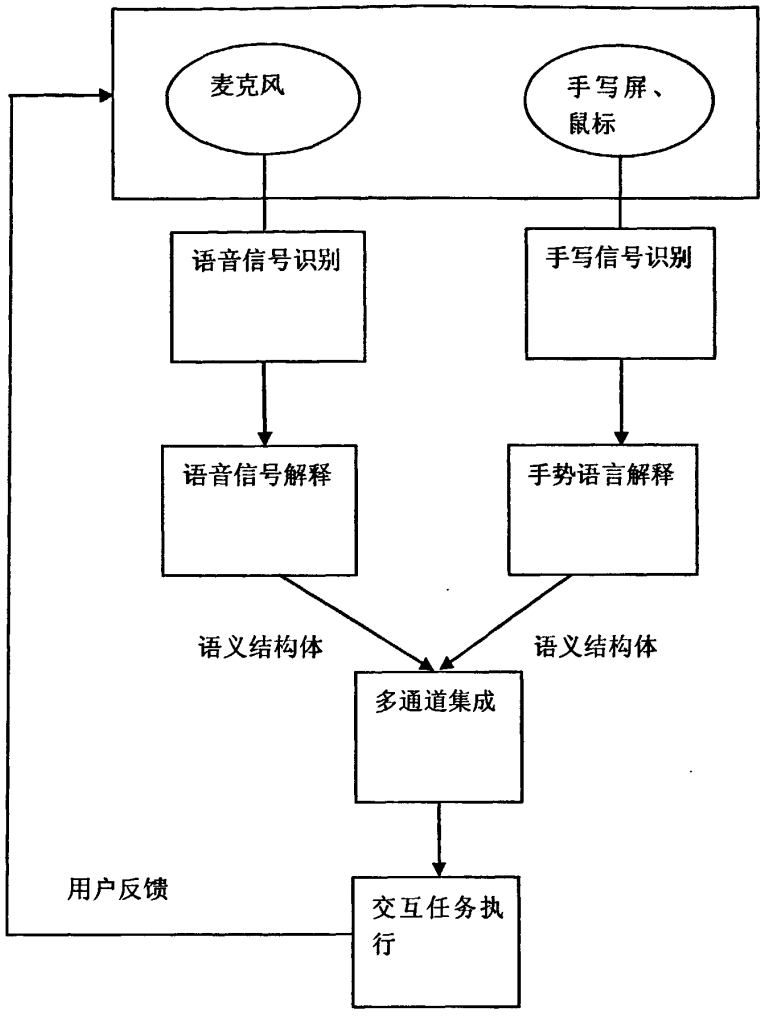


图 5.1 多通道信息处理流程

5.5 本章小结

本章首先介绍了交互原语的相关概念，然后通过一个实例描述了在完成交互任务过程中，信息的融合过程，最后介绍了对多个通道交互信息的并行处理。

第六章 几何学习系统的设计和实现

几何学习系统是一个面向中小学的多通道交互系统，在第4、5章我们介绍了该系统中的关键技术，以下主要对设计和实现中其他的工作做一些介绍。

6.1 几何学习系统的需求分析

需求分析是一个对用户意图不断进行揭示和判断的过程，要对可行性分析确定的系统目标和功能做进一步的详细论述，确定用户的要求是什么。但目前很多捕获需求的方法侧重于对功能进行建模，而对用户的心理需求并没有引起足够的重视，最终导致软件可用性不好。

6.1.1 以用户为中心的场景设计

场景可以简单的理解为关于人及人的活动的故事。基于场景的设计是一种以场景为核心的系统设计方法，它使用场景作为核心的描述并贯穿系统开发的整个生命周期[44]。它将设计的焦点从定义系统的操作转变到描述什么人将使用该系统去完成工作任务及其他活动[45]。这对用户提取用户界面的需求提供了很好的方式。

根据这种方法，我们先给出传统用纸笔和黑板学习时的观察场景：

1. 用户习惯有较大的工作区域进行绘制图形和书写文字。
2. 教师授课时，经常一边在黑板上书写板书，一边说出讲课的内容，口和手两个通道并行工作。
3. 当用户对绘制的图形不满意时，往往需要重新绘制，不能对已绘制的图形进行编辑，效率较低。
4. 教学时板书的内容得不到保存，每次教师都需要重新绘制。
5. 用户使用的学习软件需要在菜单或按钮中寻找适合的命令，通过鼠标精确的定位来完成交互。

6.1.2 分析场景

通过以上两个场景的描述，提取用户的交互信息，建立分析场景。

1. 交互的主要通道：

   黑板，纸张，笔，语音，投影仪。

2. 交互主要信息：

   手绘图形，文字，投影图片，语音信息。

3. 用户意图：

   板书与语音讲解并行工作，将信息输出，两条通道互不干扰。

根据这种方法我们发现，在中小学生学习这个场景中，人们工作的对象是黑板，纸，笔。用户将注意力集中在任务上，而纸和笔的工作方式可以让用户忘记交互的通道。当前无处不在的计算所倡导的，也正是使用户在不知不觉中享受计算服务。如果要满足在学习环境中的交互要求，显然纸和笔的交互方式是最适合的。其次，这种传统绘图方式的特点就是随意、不精确。绘制过程自然高效，但如果想要手工绘制出规整的图形，就要借助直尺，三角板等工具，效率随之降低。

6.2 系统界面设计

系统的用户界面是计算机与其使用者的对话接口，界面设计和交互设计是以用户为中心设计的两个重要内容。因为所有产品最终都是为人服务的，所以我们只有在设计当中考虑人的因素，才能设计出有较好可用性的界面。

6.2.1 界面隐喻

隐喻是指现实世界中已经存在的事物对某个对象进行比拟描述。界面隐喻的描述对象是用户界面，利用物理世界中的某事物的组织形式和交互方式与用户界面某些控制相似的特点，使人们把对这些事物的知识运用到用户界面上来，减轻了用户的认知负担。

图形用户界面的成功直接来自于隐喻的运用。比如文件操作就来自于桌面办公文件的操作。采用界面隐喻，增加了用户与应用的初始熟悉度。

界面隐喻是开发概念模型的另一种方法。比如某物理实体要和开发的概念模型的特征相似，这时我们可以利用界面隐喻。

6.2.2 PIBG 交互范式

目前Microsoft的Windows和Unix系统中的Motif窗口系统采用的是WIMP（Window，Icon, Menu, Pointing, Device）的交互风格[46]。其中：

1. 窗口：是一个矩形区域，表示了一个逻辑对话线索。

2. 图标：表示一个挂起的对话线索（即关闭的窗口），用于节省屏幕空间并提示用户以后可以恢复该对话。

3. 菜单：表示系统可以执行的命令或服务的选项，这样的可视提示辅助用户回忆与操作相关的信息。

4. 指点设备（如鼠标）：提供了用户输入信息的能力，WIMP 所需的交互风格很大程度上依赖于指点和选择交互对象（如图标）。

WIMP 所采用的多窗口策略，为用户提供了多个对话线索，可以使用户方便的在各个窗口之间来回切换，避免了以前线性的工作方式。特别适合用在交互密集型的任务。

但是 WIMP 也存在着固有的缺点。在 WIMP 界面占统治地位的几十年中，随着计算机硬件设备的进步和软件技术的发展，WIMP 界面的缺点逐渐地体现出来。对于 WIMP 界面而言，终究是局限在 Desktop 范式之上的，用它来进行文档的处理等工作非常适合，但对于其他许多应用而言，WIMP 界面并不适合。从 90 年代初开始，研究者们将研究的焦点重新集中到下一代用户界面的研究上[47]。

中科院软件所通过对 POST-WIMP 交互特征的研究，提出了 PIBG 交互范式。PIBG 范式摒弃了 WIMP 范式的繁琐，采用简洁的设计风格，使用 Pen/Paper 隐喻，模拟人们数千年来形成并熟悉的纸笔交互环境来构造笔式用户界面。从信息呈现到交互方式都有了根本性的改变，P，I，B，G 分别与 WIMP 范式的 W，I，M，P 相对应。在 PIBG 范式中，承载应用信息的交互组件由窗口（window）变为物理对象（physical object），P 是这一类交互组件的统称，主要包括 Paper 和 Frame 两类交互组件。I，B 表示此范式中与具体语义无关的直接操作组件，I 是 Icon，B 是 Button。在此范式中摒弃了 Menu 类的交互组件，尽量多地使用 Icon 和 Button，这样可以大大增加直接操作在整个交互方式中的比例，提高系统的操作效率。G 是 Gesture，是此范式中所采用的主要的交互方式。与 WIMP

交互方式比较，用户的交互动作由鼠标的点击变为笔的 Gesture[48]。

PIBG 范式并没有在各个方面完全替代 WIMP 范式，它保留了 Icon，Button 等直接操作组件。Gesture 是 PIBG 范式中用户同界面交互的主要方式，用户通过 Gesture 来对纸、框或其它组件以及框中特定内容进行处理。基于 Gesture 的交互方式同样模仿了人们千百年来在纸上用笔进行交互的方式，可以减轻用户对交互方式的认知负担，减少用户的训练时间，提高操作效率。这种方式与 WIMP 范式下利用菜单、按钮等交互组件的方式不同，不需要用户关注任务的执行过程，避免了所关注的焦点发生变化，从而能减轻用户的认知负担，提高操作效率[49]。

基于以上优点，中科院软件所开发了笔输入开发平台——PIBG Toolkit。本系统就是采用该平台为底层支撑平台。PIBG Toolkit 中包含了纸，框等多种交互组件，定义了纸，框和内容三个层次之间的静态结构和动态机制。开发者开发笔式交互系统时，可以用 PIBG Toolkit 来建立整体的软件框架和交互机制，并有选择的在系统中添加交互组件。本系统选用了 Toolkit 中提供的纸作为交互组件。

6.2.3 多通道交互界面设计

考虑到本系统的主要用户是中小学生，所以采用以用户为中心的场景设计方式设计系统界面。

由于中小学生在学习数学知识时，时常会感到枯燥乏味，精力不集中。所以在设计界面的时候多采用明快清新的色调。系统界面如图 6.1。界面中有意识的添加了一些有趣的设计元素，比如可爱的卡通形象。按钮也都采用卡通的设计风格，让人看起来心情愉悦，由此来缓解学生在学习当中的压力，集中学生的注意力。

系统主要的绘图区域，用一个记事本的图片做背景。用这种界面隐喻的方式可以清楚的表示用户在此区域绘图，一目了然的表达了系统的功能，减轻了用户的认知负荷。

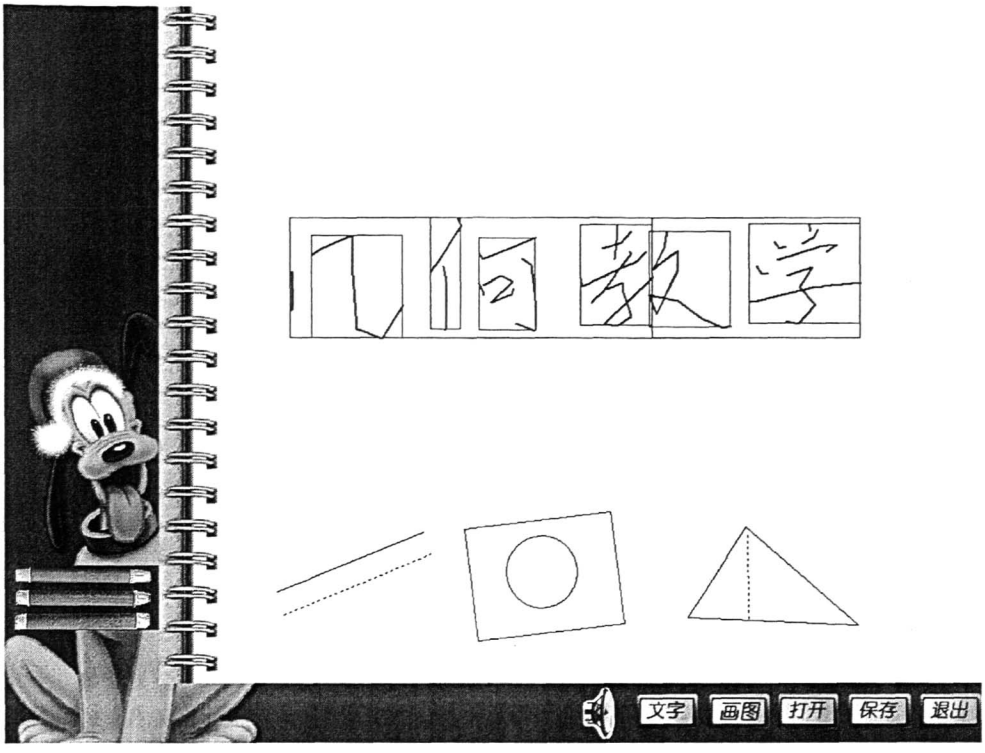


图 6.1 几何学习系统界面

按钮是笔式用户界面中命令的直观呈现，代表系统提供的某种功能。在本系统中，按钮也用特定的位图表示，如用户对笔迹颜色的选择可以通过点击界面左侧的笔状按钮实现。按钮设计成不同颜色的水笔形状，这种隐喻的方式直观的表示了此按钮改变绘图颜色的功能。用户想要改变绘图的颜色，可以直接点击相应的“水笔”。

除了纸和笔的交互，系统的语音交互界面相对简单。

当教师授课时通过麦克风输入一句话，系统会实时的把这句话显示在系统界面的顶部位置，这样听课的学生能够看清刚才没有听清的内容。当教师再说下一句话，系统会删除刚才识别出的文字，然后再显示当前识别出来的结果。由于是语音交互，所以没有复杂的界面。这种基于学习场景设计出的多通道交互界面使得学生对几何的学习更加简单方便高效。

6.3 系统总体结构

6.3.1 系统说明

中小学生与计算机交互时相对随意、不精确并且追求自然，而简单自然高效的多通道交互能够有效的解决这个问题。学生和教师在使用本系统时，可以使用笔任意的勾画各种几何图形。用户可一笔也可多笔勾画出一个简单的几何图形，如三角形、矩形，系统会在第一笔落笔的瞬间启动计时器，当用户最后一笔抬起后，系统会进行识别。系统将中小学生画的草图识别成规则的几何图形，可以加深图形结构在学生脑海中的印象。该系统当前的版本可以识别圆，直线，矩形等五种图形，图形的线条可以是实线也可选择虚线。如学生需要在图形上做辅助线就可以使用虚线来绘制直线，绘制有误还可删除，这与真实的学习过程一致。鉴于系统图形的识别率并不是百分之百，当图形系统识别不出来时会发出提示，并按用户绘制的形状显示。

当用户需要编辑绘制出的图形时，可以使用特定的命令手势。系统为编辑图形设计了三种命令手势：删除、撤销和移动。此外用户可以靠单击一下纸来进行文字输入和绘制图形两大功能的切换。系统的文字输入功能采用的是保存文字笔迹的方法显示。以这种方式显示亲切自然，符合当今软件发展的趋势。

系统的另一项主要功能就是语音交互。语音相对与其他交互方式更加自然，认知负荷相对较低，不需要一直占据用户的注意力，因此非常适合于人们在利用其他通道进行工作时进行协同操作。

绘图时，用户可以选择红黄绿三种颜色。用户即可以点击左边水笔形状的按钮，也可以用语音功能来实现。如用户可以对麦克风说出“红色”，系统会将画笔的颜色改为红色。

语音识别在本系统中的另一大功能就是在教师使用时，实时的显示教师讲课时所说的话。教师当输入一句话或一个词，系统会在停顿的时候进行识别并显示，这样哪怕坐在最后一排的学生没有听清刚才所说的句子，也可以看清这句话。另外显示识别的语句，也是为语音的交互提供一个可视的界面，对语音识别的结果进行反馈。

6.3.2 系统结构框架

此系统完成一次操作，主要由交互信息整合、任务管理以及应用程序执行三部分完成，如图 6.2。

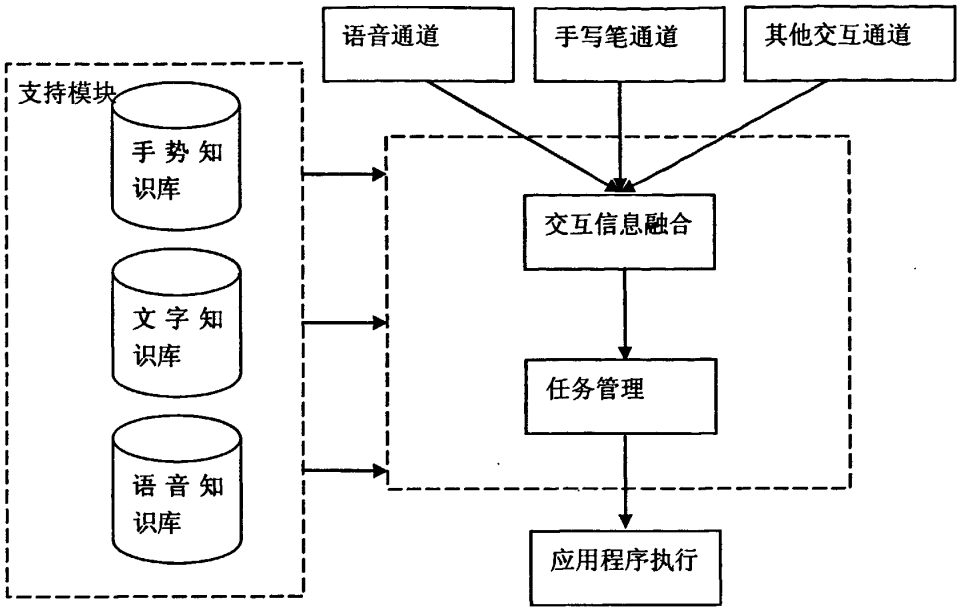


图 6.2 系统体系结构

交互信息融合负责处理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、语音和鼠标键盘设备。系统接收到交互信息整合后发来的交互任务，按照不同任务的具体要求对交互数据进行相应的语法和语义反馈 。按照通道的划分，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。

6.4 语音识别的工作流程

6.4.1 相关软件介绍

Microsoft 的语音软件开发包（Microsoft Speech SDK）是用于开发语音软件的常用工具，它包括了语音识别和语音合成引擎的最新版本，开发者主要依靠它的语音应用程序接口（Speech Application Programming Interface, SAPI 5.1）来开发软件。SAPI 的 API（Application Programming Interface, API），以 COM 组件的形式提供，程序员不用去了解复杂的语音识别技术就可以开发语音应用

程序，而且这样还极大的减少了编写语音识别和语音合成应用程序所需的代码，使得语音技术更加容易使用。

6.4.2 语音识别的工作流程

语音识别过程需要用到两个重要的接口：语音识别上下文接口（ISpRecoContext）和语音识别引擎（ISpRecognizer）。上下文接口提供了为请求语音识别的事件接受通知消息的基本载体。

首先需要进行语音识别模式初始化，需要分三步来实现：

1. 创建一个语音识别引擎

调用 IspRecognizer 接口的 CoCreateInstance 方法。用 CLSID_SpSharedRecognizer 参数创建一个共享引擎，用 CLSID_SpInprocrecoInstance 参数创建一个进程内引擎。

2. 为识别引擎创建一个上下文接口

为了创建共享 IspRecognizer 的 IspRecoContext 接口，应用程序需指定参数为组件的 CLSID_SpSharedRecoContext 并调用 COM 的 CoCreateInstance 函数即可。这时，SAPI 将设置音频输入流为 SAPI 的默认音频输入流。

3. 为应用程序感兴趣的事件设置通知消息

IspRecognizer 也是一种 IspEventSource 接口，是 IspNotifySource 接口的一种。因此，应用程序能够从其 IspRecoContext 接口中调用 IspNotifySource 的方法来指定 IspRecoContext 所需的事件向何处通知。然后调用 ISpEventSource::SetInterest 方法可以设定什么样的事件需要被通知。最重要的事件是 SPEI_RECOGNITION，它显示了 IspRecognizer 已从 IspRecoContext 中识别了一些语音。

4. 为应用程序创建，装载并激活一个语音识别语法接口

语音识别语法接口（IspRecogGrammar）指示需要的语音识别类型，即连续语音识别（dictation）或者控制命令识别（command and control）。

应用程序首先使用 ISpRecoContext：：CreateGrammar 创建一个 ISpRecogGrammar 接口，然后装载合适的语法。调用 ISpRecoGrammar::LoadDictation 装载 dictation 语法，使用某个 ISpRecoGrammar：：

LoadCmdxxmethods 方法装载 command and control 语法。

最后，激活这些语法使识别引擎开始工作，应用程序调用 ISpRecoGrammar:: SetDictationState 设置口述状态。调用 ISpRecoGrammar:: SetRuleState 或 ISpRecoGrammar:: SetRuleIdState 激活 command and control 语法。

语音识别模式初始化后，识别引擎便开始工作，如果识别语法里期望的内容被识别出来，识别引擎便发出 SPEI_RECOGNITION 消息。因此，应用程序需要响应 SPEI_RECOGNITION 消息。识别的结果都包含在语音识别结果接口(IspRecoResult)里，IspRecoResult 里同时还封装了用于提取结果的方法。应用程序可以通过调用这些方法从 IspRecoResult 的数据结构里得到识别的词语和句子，识别的语法规则名字和 ID 号，以及对应的命令 [50]。

6.5 原型系统使用分析

使用该学习系统时，学生需要一台手写屏，一台 PC 以及麦克风，教师授课时可以使用投影仪。当然如果没有手写屏和麦克风，用户仅用鼠标也完全可以完成操作。

开始使用该系统时，系统默认情况是开启语音识别功能。当不需要语音识别时，用户可以点击界面下方的喇叭按钮，关闭该功能。

以下以教师的身份，通过讲解“直线和圆的位置关系”，来演示该系统的使用。

用户在绘图区域可以简单的勾画圆和直线，系统会识别出用户想绘制的图形并显示。同时用户可以使用“文字切换”手势“丨”，或点击按钮“文字”切换到板书状态来写文字。这个操作也可以通过向麦克风说出语音命令：“板书”完成，如图 6.3。

绘制过程中，用户可以用语音命令“蓝色”，“红色”来改变笔的颜色，如图 6.4。用户也可以把直线拖动到与圆相交或相切的位置，通过动态的变化，形象的演示直线和圆的位置关系。

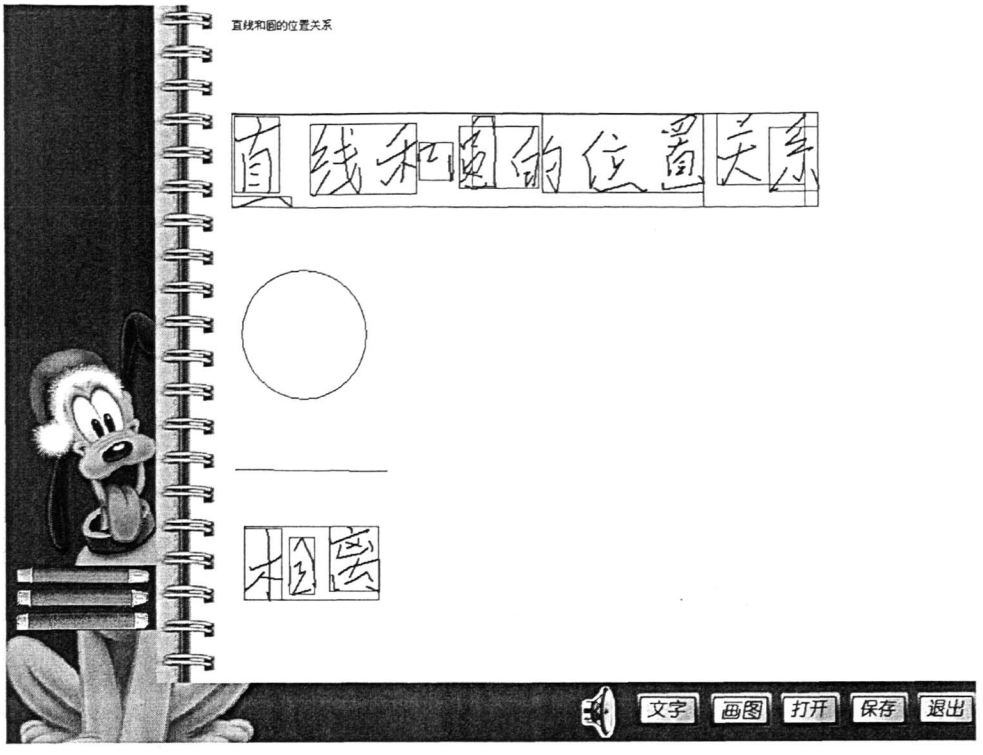


图 6.3 系统实例 1

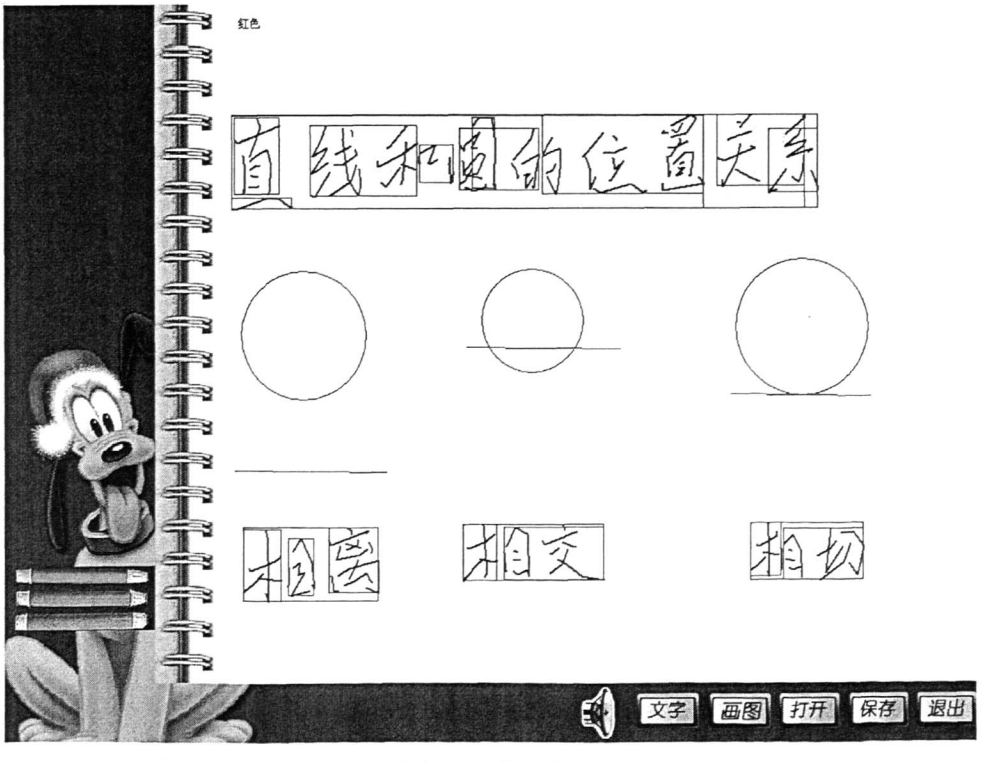


图 6.4 系统实例 2

最后用户可以发出语音命令“退出”来退出该系统。

用户可以在学习时绘制图形，也可以打开课前绘制好的课件学习。用户可以点击“打开”按钮，也可以对麦克风说命令“打开”，这两个动作的执行效果相同。

6.6 本章小结

本章通过对几何学习系统的设计和实现，对多通道系统在学习领域的应用做了新的尝试和探索。鉴于用户的特殊性，系统在遵循以用户为中心的原则进行设计时，做了相当大的工作。本章首先介绍了系统需求的获取。由于不同于传统的交互方式，本章又重点介绍了 PIBG 交互范式和原型系统的界面特点。随后对该系统的总体结构做了简介。最后详细介绍了该系统的使用。

第七章 总结和展望

7.1 论文工作总结

本文首先对人机交互中手写识别和语音识别两种技术做了介绍，然后对多通道交互中的关键技术做了简介，随后详述了系统开发过程中的两项重要工作：手势识别算法和信息融合策略。最后介绍了面向中小学的几何学习系统的研究和实现。

将自然高效的语音交互技术、笔交互与图形识别算法结合在一起，开发的面向中小学的几何学习系统，在功能上弥补了以往电子白板的缺点。在几何图形识别方面采用一种与笔画顺序、笔画数目、方向无关的识别算法，具有识别率高、符合用户手绘习惯的优点。

在功能上，该系统弥补了同类教学软件无法支持板书，图形识别率不高等缺点，并增加了语音交互，提供了更自然高效的交互方式。当前有很多研究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，无法捕获用户输入的全部信息；有些交互技术对中小学生并不适用，年龄较小的用户难以掌握系统的使用。与同类软件系统相比，本系统在一定程度上克服了当前普遍存在的儿童难以使用的问题。无论是中小学生还是教师，都可以通过笔和语音快速、便捷的与系统交互，简单随意的勾画，系统就能够高效的识别出想要得到的图形。虽然当前系统的版本还有一些问题存在，相信该系统配以投影仪，手写板也能大大提高教学质量，减轻教师负担，具有广阔的应用前景。

7.2 存在的问题及展望

在下一步的工作中，还要对系统做如下完善：

1. 笔输入时命令手势较少，还不能满足教学和学习中的操作需要，以后会进一步补充实现。

2. 语音识别率太低，影响了许多功能的开发，也降低了使用效率。下一版

本将使用识别率相对较高的语音库解决此问题。

3. 增加语音的反馈功能，通过发声、加入音乐等方法，进一步提高系统的趣味性。

4. 语音输入时能识别的命令较少，如能增加更多命令用于对图形编辑操作，将大大增强该系统的实用性。

总之，多通道将是未来人机交互的技术特征。它主要体现在：

未来计算机发展的“隐身化”和“微型化”使得界面不局限于屏幕，而是和更多的交互设备联系起来，其中“嵌入式”系统的推广将更加推进这方面的发展[51]。

自然、高效将是未来用户界面的感知特征，尤其对于正常人群之外的群体，比如残疾人、文盲等。

个性化定制将是未来用户界面的功能特征之一。未来用户界面将逐步做到“计算机适应人”，从追求“容易实现”到“容易学习和容易使用”，将明显突出用户本身的兴趣和爱好。

表现形式的多样化将是未来用户界面的应用特征。由于因特网、无线设备以及移动计算技术的发展，人类已经进入因特网分布计算的新纪元，用户范围更加广泛，使用要求也更加多样化，用户界面的发展必须体现这种要求。

语音识别和指点方式的结合将是未来用户界面的主要形式。当前语音识别技术和具有触觉反馈的笔输入技术日趋成熟，视觉是人们接受信息的主要通道，语音、笔交互、手势是人们进行交互的主要手段。随着多通道交互技术的发展，除了对多个通道交互的结合之外，多个通道在特征层次或者语义层次上的融合越来越重要[52][53]，这也是我们下一步研究的目标。

参考文献

[1] 付永刚，戴国忠，蒋成高，藤东兴. 支持笔输入的虚拟家具设计系统. 计算机辅助设计与图形学报. 2002(9):877-879

[2] Myers B, Pausch R, Pausch R. Past present and future of user interface software tools. ACM Transactions on Computer-Human Interaction. 2000, 7(1):1-28

[3] 俞美英. 计算机用户界面及发展. 安徽工学院报. 1997(4): 24-27

[4] Wilcox L, Schilit B, Sawhney N. Dynomite: A Dynamically Organized Ink and Audio Notebook. Proceedings of CHI'97, 1997.186-193

[5] 田丰，牟书，戴国忠，王宏安. Post－WIMP 环境下笔式交互范式的研究. 计算机学报，2004，27(7):977-984

[6] 田丰. Post－WIMP 软件界面研究. 中国科学院软件研究所博士学位论文，2003 年 6 月 1-4 页

[7] Tolba O, Dorsey J, McMilan L. A projective drawing system. In: Hughes, J.F., Sedquin, C.H., eds. Proceedings of the 2001 ACM Symposium on Interactive 3D Graphics. New York: ACM Press, 2001.25-34.

[8] Tolba O, Dosey 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

[9] Vaucelle C, Jeban T. Dolltalk: a computational toy to enhance children's creativity. In: Burleson, W., Selker, T., eds. ACM CHI 2002. 776-777.

[10] Dong Shi-hai, Wang Jian, Dai Guo-zhong. Human-Computer Interaction and Multimodal User Interface. 1st ed., Beijing: Science Press, 1999(in Chinese).

[11] 宋学义. 一个基于笔的概念图编辑器的设计与实现. 西北大学硕士学位论文，2007 年 6 月 5 页

[12] Norman D. Concentions and Design. ACM Interactions Magazine, May/June 1999. 38-42.

[13] Dix A, Finlay J, Abowd D G, Beale R. Human-Computer Interaction. Third Edition. 北京：电子工业出版社.2006.30-43

[14] Weiser. The Computer for the 21stCentury. Scientific American，1991.66-75.

[15] 秦真，饶文碧. 多通道交互及其信息融合技术的研究. 武汉理工大学硕士学位论文， 2006 年 4 月，8 页

[16] Gross M D, Do E Y. Ambiguous intentions: a Paper-like interface for Creative Design. UIST'96 Seattle Washington USA, 1996. 25-34

[17] Stahovech T F, SketchIT: a Sketch Interpretation Tool for Conceptual Mechanical Design. Massachusetts Avenue, Cambridge, MA: Massachusetts Institute of Technology, 2002

[18] Moran T P, Chiu P, Melle WV. Pen-Based Interaction Techniques for Organizing Material on Electronic Whiteboard. UIST, 1997.45-54

[19] Pedersen E R, McCall K, Moran T P, Halasz F G. Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings, In Proceedings of the ACM INTERCHI'93 Conference on Human in Computing Systems, 1993.391-398

[20] Moran T P, Chiu P. et al. Implicit structures for pen-based systems within a freeform interaction paradigm. proceedings of CHI'95, 1995.487-494.

[21] 雷葆华，和应民. 语音用户界面平台的设计与评估. 哈尔滨工程大学. 2002 年 6 月， 25 页

[22] 蒋宇全，罗军，林应明，董士海. 基于任务的多通道整合设计与实例. 计算机学报，1998(9):860-864

[23] 吴玲达，王晖. 多媒体人机交互技术. 长沙：国防科技大学出版社. 1999.

[24] Shneiderman B. The Limits of Speech Recognition: Understanding acoustic memory and appreciating prosody. Communications of the ACM, 2000(9):63-65

[25] Rudnicky A. Creating natural dialogs in the Carnegie Mellon Communicator system. In: Proceedings of the 6th European Conference on Speech Communication and Technology(EuroSpeech). Budapest, Hungary, 1999. 1531-1534

[26] Rudnichy A I. Task and Domain Specific Modeling in the Carnegie Mellon

Communicator System. Proceedings of ICSLP 2000(Beijing, China) Paper G4-01.

[27] Weinschenk S, Barker D T. Designing Effective Speech Interfaces. Inc, 2000. 100-106P, 110-120P, 210-239P

[28] 方志刚. 视线跟踪技术及其在多通道界面中的应用. 系统工程与电子技术， 1999(6):46-49

[29] 任海兵，竹远新，徐光佑. 基于视觉手势识别的研究综述. 电子学报，2000(2):118-121.

[30] Alvarado C J, Davis R. Resolving ambiguities to create a natural sketch based interface. Proceedings of UCAI-2001, August 2001

[31] Myers B A. New Models for Effective User Interface Software Development. IEEE Teansactions on Sofrware Engineering, 1997.23(6) :347-365.

[32] Landay J A, Myers B A. Sketching interfaces: Toward More Human Interface Design. IEEE Computer, Vol.34, no.3, March, 56-64.

[33] Lin J, Newman M, Hong J, Landay J A. DENIM: Finding a Tighter Fit Between Tools and Practice for Web Site Design. CHI Letters: Human Factors in Computing Systems, CHI'2000,2000.2(1):510-517.

[34] Hong J I, Landay J A. SATIN: A Toolkit for Informal Ink-based Applications. In Proceedings of UIST'00 Symposium on User Interface Software and Technology(Nov.6-8, SanDiego, CA), ACM, 2000.63-72

[35] Mynatt E D, Igarashi T, Edwards W K, LaMarca A.Flatland:new dimensons in office whiteboards. In Proceedings of CHI'99 Human Factors in Computing Systems(May 15-20, Pittsburgh, PA),ACM, 1999.346-353.

[36] Forsberg A, Dieterich M, Zeleznik R. The Music Notepad, In Proceedings of the ACM Symposium on User Interface Software and Technology:UIST'98, 3-10. San Francisco, CA.Nov.6-8.

[37] 吕争，吴名慧. 基于语音的人机交互界面的研究与实现. 荆门职业技术学院学报， 2007,22(6):15-16

[38] Jennifer Preece Yvonne Rogers Helen Sharp, 交互设计—超越人机交互.北

京：电子工业出版社. 2003.1-48

[39] 李杰，田丰，王维信. 面向儿童的多通道交互系统. 软件学报， 2002,13 (9) :1846-1851.

[40] 敖翔，戴国忠. 数字笔迹的结构分析与识别. 北京：中国科学院研究生院，2006 年 6 月 78 页

[41] 孙正兴，冯桂焕，周若鸿. 基于草图的人机交互技术研究与发展. 计算机辅助设计与图形学报，2005.9.17(9):46-48

[42] 马翠霞. 支持概念设计的手势描述和草图设计系统的研究. 中国科学院软件研究所博士学位论文，2003 年 6 月 10 页

[43] 普建涛，陈文广，王衡，董士海. 多通道用户界面关键技术和未来发展趋势研究. 计算机研究和发展，2001.3.8(6):12-13

[44] Carroll J M. Scenarios and design cognition. In: Proceedings of the APCHI 2002 User Interaction Technology in the 21st Century, Beijing. 2002,23-46

[45] Raskin J, The Human Interface, Addison Wesley Publishing Company, 2000; 北京：机械工业出版社（影印版），2002.24-41

[46] 华庆一. 以用户为中心的系统分析、建模与设计过程研究. 西北大学博士学位论文，2006 年 6 月，32 页

[47] Igarashi, Takeo, Matsuoka, Satosbi, 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.

[48] 刘婷婷. 支持概念设计的二维 CAD 系统的设计与实现，西北大学硕士学位论文，2007 年 6 月 5 页

[49] 田丰，秦严严，王晓春，翱翔，王宏安，戴国忠. PIBG Toolkit：一个笔式界面工具箱的分析与设计. 计算机学报，Vol.28，No.6，2005

[50] 李禹材. 通用语音控制命令识别 COM 组件的研究与开发. 四川师范大学硕士学位论文，2004 年 1 月，36 页

[51] 朱军，张高，华庆一，戴国忠. 交互式用户界面的形式化描述和性质验证. 软件学报，1999，10(11):4-6

[52] Oviatt S, DeAngeli, A, Kuhn K. Interation and synchronization of input

modes during multimodal human-computer interaction. Proceedings of the CHI'97, ACM Press, NY,415-422

[53] Grasso A M, Ebert D S, and Finin T W. The Integrality of Speech in Multimodal Interfaces. ACM Transactions on Computer-Human Interaction, 1998, 5(4):303-32

致 谢

首先衷心感谢我的导师华庆一教授。本文是在华老师的支持、鼓励和精心指导下完成的，论文完成过程中的每一步进展都凝聚着导师的心血。华老师学术思想的活跃性，给了我无限的启迪，是我灵感的源泉。他丰富的专业知识、严谨的治学态度、广阔的视野和深刻的思想方法，减少了我在学习上的困难，使我的思维更加开阔，知识更加全面。尤其是在论文的选题、撰写过程中，华老师都给了我极大关注和指导，论文的完善和定稿华老师更是付出了大量的精力，在此谨向华老师表示由衷地感谢！

感谢三年来朝夕相处的同学们，我们在一起生活学习，相辅相伴走过了三年的研究生学习生涯，你们给了我最无私的帮助，和你们在一起的日子我将终生难忘。我将永远珍视这珍贵的友谊。在此对人机交互实验室的师兄师姐和师弟师妹们表示衷心的感谢。

感谢我的爸爸妈妈，对我在校学习的支持，对我无微不至的关怀和包容！

攻读硕士期间完成的论文

[1]. 王爽, 华庆一. WEB 系统维护中逆向工程研究与实现. 计算机技术与发展. 已录用。