

通道界面, 模型, 人机系统,

多通道界面模型与关键技术^{*)}

多媒体技术

97, 24(1)

李茂贞 戴国忠

董士海

(中国科学院软件研究所 北京 100080)(北京大学计算机系 北京 100087)

TP391

A

摘要 Based on new interaction techniques, such as eye-movement tracking, speech recognizing, gesture input and sensation feedback, multimodal interface uses more than one sensation and action modalities to interact with users, thus promoting the nature of human-computer interaction. In addition, integrating the accurate/inaccurate information of multimodalities, if can expand the bandwidth of user input and improve the efficiency of interaction. Firstly, this paper presents the motive of multimodal interface research, and then gives a difference between the multimodal interface and multimedia interface based on the interface model we present here. Finally, it raises some key technologies involved in the multimodal interface and potential goals are explored.

TP11

1-4

关键词 Human-computer interaction, Multimodal interface, Multimodalities integration.

一、引言

随着计算机科学技术的飞速发展,计算机日益进入普通用户的工作和生活中。面对数量剧增的普通用户,人机交互的自然性在系统设计中的地位越来越重要。针对这样的情况,Norman^[1]等人提出了以用户为中心的系统设计,用户利用自然技能无需专门的计算机系统操作技能就可以自如地使用计算机,体现人机交互的自然性。虚拟现实正致力于这方面的研究。利用头盔、数据手套等交互手段,用户在操作系统时有身临其境的感觉。另一方面,多媒体技术的迅猛发展,使得计算机到用户的输出不再是单一格式的字符流,而是文本、图形、图象、声音等多种媒体的集成,这样计算机到用户的通讯带宽大大提高。但是,目前的人机交互由于一直受到基于键盘、鼠标的窄带宽输入界面的限制,使得用户到计算机的通讯带宽仍停滞不前,人机交互中输入与输出效率变得越来越不平衡。此外,在传统的人机交互中,用户只能利用键盘/鼠标一种输入设备来指定一个或一系列完全确定的命令或参数,因此,用户输入以串行性和精确性为特征,在许多场合不必要地增加了用户的作业负荷,降低了交互效率,破坏了交互的自然性。为了提高人机交互的自然性、高效性,必须

打破常规,采用多通道用户界面。

多通道用户界面(Multi-Modal user Interface, MMI)中“通道(modality)”一词源于心理学,其词源为“mode(方式或模式)”。在讨论视觉、嗅觉、触觉等感觉的“方式(非内容)”时,或者在涉及多种感觉之间的关系或讨论一种感觉不同于其它感觉的特点时,心理学中使用“感觉通道”一词。本文把“通道”视为“方式”的同义词,不仅表示用户的“感觉方式”,也指用户的“动作方式”。通道与实际的硬件设备有关,有几种交互设备,就对应几种交互通道。多通道用户界面采用视线跟踪、语音识别、手势输入等新的交互技术,允许用户利用多个通道以自然、并行、协作的方式进行交互,通过整合多个通道精确/非精确输入,捕捉用户的交互意向,提高人机交互的自然性和高效性。

八十年代后期以来,多通道交互技术的研究作为人机交互的一个新领域在欧美越来越受到重视。目前正在美国进行的相关研究项目主要有MIT媒体实验室的“多通道自然对话”,CMU交互系统实验室(ISL)的INTERACT项目,NRL Intelligent M4系统研究组;在欧洲信息技术研究战略规划(ESPRIT II)的Amodeus项目中有大量关于多通道人机交互的理论和系统研究。在ESPRIT于III中,正在

^{*)}受国家自然科学基金重点项目支持,项目批准号为69433020。李茂贞 博士生,研究方向为人机交互、软件工程;戴国忠 研究员,博士生导师,研究方向为计算机图形学;董士海 教授,研究方向为计算机图形学。

进行 MIAMI(Multimodal Integration for Advanced Multimedia Interfaces)的项目,其研究领域包括多媒体和高级多通道人机界面两部分,语言内容是通过视觉、听觉、以及触觉、手势来访问、表示和产生多媒体信息的多通道交互有关的各个方面。在法国 IMAG 的 Coutaz 和 Nigay 设计的系统 MATIS^[2] (Multimodal Airline Travel Information System) 中,用户可以利用键盘、鼠标、话筒或其组合方式查询航班信息,体现了多通道用户界面人机交互的自然和高效性。

由北京大学计算机系、中国科学院软件研究所和杭州大学心理系共同承担的“多通道用户界面研究(Multimodal User Interface Research, MUIR)”是国家自然科学基金重点项目,主要探索以视线跟踪为核心的多通道人机交互的理论和新技术,内容包括多通道用户界面的输入技术和通道信息整合方法,适合于多通道界面的用户模型和描述方法,支持多通道交互的用户界面模型、设计原则、评价体系和开发环境以及在虚拟座舱和 CAD 方面的应用实例。

二、多媒体界面模型^[3]与多通道界面模型

如图 1 和图 2 所示,用户界面在用户和系统功能之间起中介作用,通过界面交互命令向应用系统提交用户任务,系统完成任务后反馈信息由界面信息表示部件表现出来。多媒体界面和多通道界面的界面信息表示可以有文本、图形、图象、声音等多种媒体。

从图 1 可以看到,目前的多媒体界面的界面交互命令只接收来自输入通道 M1(M1 可以是键盘或鼠标)的单通道、串行、精确的输入,与传统的用户界面相比,多媒体界面只是在界面信息表现方式上进行了改进,采用了多种媒体,即计算机到用户的通讯带宽提高,但是,用户到计算机的通讯带宽却未提高,人机交互效率仍然很低。相比之下,多通道用户界面在界面信息仍表现为多媒体的基础上,可以提高人机交互效率。

从图 2 看到,多通道界面中内嵌了通道整合部件,可以接收来自 n 个输入通道(M1, M2, ..., Mn)的信息流,多个通道通过整合部件可以并行、互补协作,从而加大了人机交互的通讯带宽,提高了交互的效率。另外,多通道用户界面也支持非精确信息的输

入(如语音和手势),提高了人机交互的自然性。由此可知,多媒体界面对媒体信息只是存储和转发,并没有进行解释,而多通道用户界面则侧重于媒体信息的理解。不难理解,多通道用户界面是多媒体界面的进一步延伸。

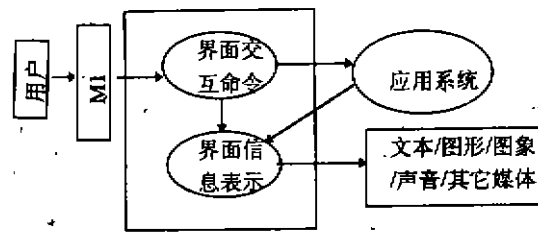


图 1 多媒体界面模型

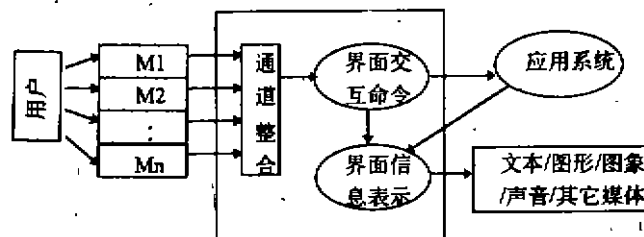


图 2 多通道界面模型

三、多通道界面研究中的关键技术

1. 用户任务分析模型

用户任务分析模型用于人机活动中用户的活动。自从 GOMS^[10]模型问世以来,已有许多种任务分析模型。任务分析模型中的一个重要问题是分解的原子层次,在多通道人机交互中这一问题变得更为复杂。以 GOMS 模型为例,它采用四种成分来描述用户行为:目标(Goal)、操作体(Operator)、方法(Method)和选择规则(Selection rule)(GOMS 是这四个成分的首字母缩写)。其中操作体指一些基本的知觉、动作或认知活动。方法是指完成某一目标的一个操作体或子方法的序列。选择规则在同一任务具有多个方法时决定选择哪一个。为准确预测用户交互的时间等性能指标,GOMS 必须分解得非常细致,而且分解的原子层次是简单的操作体。在多通道人机交互中,用户动作的分解面临多个感觉和动作通道活动的复杂组合问题,不再有简单的原子操作。与此相关,以前的任务分析模型极少考虑用户活动的并发性,因此多通道交互的任务分析模型必须适

合多通道界面的特点。从用户的角度而言,要能够表达用户利用多通道协作完成任务的特点,应能很好地分析和描述交互过程中用户的活动,体现以用户为中心的设计宗旨;从系统角度而言,要能够处理多种交互手段的并行协作使用与系统处理的关系。

2. 多通道界面图形描述方法

多通道用户界面基于多种输入通道,以用户为中心,以用户使用的自然性为宗旨,需要对用户的工作负荷、视觉反应等人素(Human Factors)作工效分析和评估,这就涉及到认知心理学知识。所有这些人机交互的特点无疑给多通道用户界面的图形描述方法提出了许多新的课题。具体来说,一个理想的多通道用户界面的图形描述方法应具备以下基本的表达能力:

(1)给设计者提供一种直观方法,使之觉得方便好用;

(2)描述多个通道的非精确输入/输出(I/O)事件。为实现多通道间的信息流整合,各个通道的 I/O 事件最好能采用统一的表示方法;

(3)方便地表示多个通道之间自然的并行、同步、选择等协作关系和约束关系;

(4)描述多通道的用户非精确输入的整合。传统人机交互的输入操作中用户每次只能使用一种输入设备来指定一个或一系列完全确定的命令或参数。在多通道界面中,可能需要整合多个输入通道的信息才能确定一个命令或动作语义,如何描述这种通道整合和应用功能映射是多通道用户界面图形描述的难点;

(5)由于涉及到多通道人机交互的自然性,所以要对多通道用户界面模型进行评估和工效分析,这就要求描述方法应该显式地反映出用户和任务分析模型的分析结果,以便进行认知心理学的评价。

3. 多通道整合

多通道界面的特点就是利用多个感觉和动作通道的并行和协作进行人机交互,多通道界面互相独立地利用多个通道并不是真正意义上的多通道界面,这并不能有效地提高人机交互的效率,如何从多个并行/串行、精确/非精确、独立/协作的输入信息流中快速捕捉用户想传达的任务信息,这就是多通道间的整合问题。在整合中需要三类信息:时间关系、语法约束和语义约束。时间关系基本上与应用无关,因为相关的多通道事件之间在时间的并行性和接近性上存在着必然的联系,只要针对特定应用总结出相应的时间参数之后便与特定的应用无关;而

语法和语义约束则是与特定应用的用户任务模型和任务结构紧密相关的。Nigay 在多通道航班信息查询系统 MATIS 中提出的整合算法^[1]基本上是基于时间的(利用了短时间、大时间和上下文整合)。由于系统在处理每个通道信息流时使用相同的时间片,实际上每个通道信息需要的时间不应相同(比如语音处理就比键盘处理所需的时间要长),再加上整合时没有充足的语法和语义信息,所以信息整合次序与用户的实际操作次序经常不一致,从而导致信息整合失败。因此多通道整合应该处理应用相关的语义,并且只有充分利用应用语义才能完成完整的信息整合。目前有关多通道整合的研究主要集中于把语音、手势和传统输入通道的整合方面^{[2][3][4]}。

4. 多通道界面软件结构

为了简化日益复杂的界面设计工作,人机交互的研究中出现了对话、应用及界面分离的原则,Seeheim^[5]模型典型地表达了这样的思想,基于这种思想实现了许多用户界面管理系统(UIMS),但是 Seeheim 模型的不足之处是支持语义反馈的能力弱。多通道主要目标是增强人机交互的语义反馈,只有在语义信息充分的情况下才有可能进行多通道整合。PAC^[6]模型和 Arch^[7]模型是两个改进的 Seeheim 模型。PAC 模型基于智能体(Agent),一个完整的交互系统可以分解为具有体系结构的多个智能体。每个 PAC 智能体都提供了一定的抽象级别,并且同一水平层次智能体其抽象级别相同。PAC 模型特点:(1)智能体是自主对象,具有一定的内部状态,可以实现并发处理;(2)智能体是具有一定功能的模块单元,一个智能体可被另一个智能体替代而不影响交互系统的其它部分,这样系统的开放性较好。它侧重于从水平角度将交互系统分解为多个 PAC 智能体,但对每个智能体的具体含义没有提供任何说明。Arch 模型侧重于从垂直角度将交互系统分解为交互部件、表示部件、对话部件、任务适配器及应用部件。其特点是利用表示部件和任务适配器把用户界面的关键部分(如对话部件)从各种具体应用功能和交互工具(如 X Window 环境)中分离出来,这样可以提高交互系统的通用性,并且减少交互系统开发的复杂性。但是,Arch 模型不支持新的交互技术如信息的并发处理、信息整合等。由此看来,PAC 模型和 Arch 模型存在互补关系,作者认为把两个模型结合起来不失为多通道界面采用的一种好的软件结构模型。

结束语 多通道用户界面利用多个输入通道的协作

进行人机交互,目的在于提高人机交互的效率,增进交互的自然性,可以将其看成是多媒体用户界面的延伸。多通道界面的研究在国外也是刚刚起步并且是交叉学科的研究,因此存在很多的问题尚待解决^[12]。作者对多通道界面中的主要问题进行了讨论并探讨了可能解决的方法,旨在起到抛砖引玉的效果。多通道界面的研究必须在很大程度上借助于心理学、认知科学、通信理论、软件工程、图形学等多方面的理论和方法^[13]。

参考文献

- [1] D. A. Norman et al., User Centered Design, Lawrence Erlbaum Associates Publ., 1986
- [2] L. Nigay, et al., MATIS, A multimodal airline travel information system, SM/WP10, ESPRIT BRA 7040 Amodeus, Feb., 1993
- [3] D. Bell, et al., General models of multimedia interaction, ERCIR Workshop Reports ERCIM-94-W003, 1994
- [4] A. Gourdoi, et al., Two Case Studies of Software Architecture for Multimodal Interactive System, VoicePaint and a Voice-enabled Graphical Notebook, In Proc. IFIP Working WG2. 7 Working Conf., Engineering for Human-Computer Interaction, Ellivuori, Aug., 1992
- [5] A. G. Hauptmann, et al., Gestures with speech for

graphic manipulation, Intl. J. Human-Computer Studies 38, 1993

- [6] S. H. Kloosterman, Design and implementation of a user-oriented speech recognition interface, the synergy of technology and human factors, Interacting with Computers, 6(1)1994
- [7] G. E. Pfaff, editor, User Interface Management System, Proc. of the Workshop on UDMS, Seeheim, Nov. 1-3, 1983. Springer, Berlin, 1985
- [8] J. Coutaz, An Object Oriented Model for Dialog Design, in Proc. Interact'87, North Holland, 1987
- [9] The UIMS Tool Developers Workshop, A Meta-model for the Runtime Architecture of an Interactive System, SIGCHI Bulletin, 24, Jan., 1992
- [10] S. K. Card, et al., The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum, 1983
- [11] L. Nigay, et al., A generic platform for addressing the multimodal challenge, in Proc. CHI'95 Human Factors in Computing System. ACM New York, Denver, 1995
- [12] John Lee, Graphics, information and multimodal HCI, ERCIR Workshop Reports ERCIM-94-W003, 1994
- [13] 罗军、董士海,多通道人机交互研究,全国第四届多媒体技术学术会议, 1995

(上接第 54 页)

是:一旦〈DF 事件〉的发生度大于等于某预先设定的阈限时,计算机就主动触发执行其后 DF 的 IF-THEN 规则。即如果〈DF 条件 1〉足够真,则执行其后的〈DF 动作 1〉,并且接着检查下一 IF-THEN 规则,直至执行完为止。在此,也可把上述事件驱动规则定义为:一旦〈DF 事件〉发生足够大后,就主动触发执行其后的 DF 规则集合(规则库),用 DFL 逻辑表示的产生式系统执行模式来解决执行相应的 DF 规则集合。

由上可见,为了 DF 地触发执行由 DF 事件驱动的规则,需有一个相应的 DF 事件监视器。它主要做两件事:1)根据要求,不时地计算各种 DF 事件的发生度。2)检查 DF 事件的发生度是否大于等于某预先设定(或系统或用户设定)的阈限,若然,则执行该规则中包含的 DF IF-THEN 规则,即根据其中 DF 条件被满足的程度,选择执行相应的 DF 动作。

DF 监视器是与其它应用软件并发执行的,动态

地监视应用软件执行过程中各种 DF 事件发生的情况,从而可以以一种统一的机制实现许多数据处理功能。

参考文献

- [1] 李凡长、李彤、沈华芬等,一种动态模糊逻辑系统,《兰州大学学报》(专刊), 1996
- [2] 李凡长、李彤、沈华芬等,动态模糊集初探,同[1]
- [3] 李凡长、丁卫群、郑家亮,动态模糊测度扩张原理,《云南大学学报》, 1997
- [4] 李凡长、丁卫群、郑家亮,动态模糊集的扩张原理,同[3]
- [5] 李凡长、丁卫群、郑家亮,动态模糊测度理论,同[3]
- [6] 郑家亮、李凡长、沈勤组,动态模糊关系,同[3]
- [7] 李凡长、沈华芬,面向主动 DF 对象的 DF 数据库系统设计方法研究,《云南民族学院学报》, 1996 No2
- [8] 李凡长著,《动态模糊逻辑及其运用》,云南大学出版社
- [9] 李凡长、沈勤组、郑家亮、沈华芬著,《动态模糊集及其应用》,云南科技出版社
- [10] 沈勤组、李凡长、沈华芬,动态模糊目标可达性研究,同[3]