

人机系统 多媒体信息 人机交互

计算机科学 1995 Vol. 22 No. 4

S-8

人机交互中的多媒体信息:希望与问题

刘正捷

(大连海事大学计算机系 大连 116024)

TP11

摘 要 Multimedia technology brings about new promise and also new challenges to the area of human-computer interaction. Based on the analysis of properties of multimedia and multimodality in human-human communication, this article raises and discusses some issues of crucial importance to human-computer interaction, such as media integration, multimedia semantics, metaphor and manipulation of time-varying media. And potential approaches to tackle the issues are explored.

关键词 Human-computer interaction, Multimedia, Multimodality, Media integration, Metaphor, Semantics, Time-varying media

一、引言

计算机应用的迅速发展越来越要求我们把整个人机系统而不仅仅是计算机系统做为考虑问题的着眼点,去研究如何提高人机系统的整体效率,进而达到提高人的工作效率这一根本目的。决定人机系统效率的一个重要因素是系统部件之间即人与计算机之间的通信。多少年来,人与计算机之间的通信手段一直局限于文本方式。这与人在自然形态中所使用的多种信息媒体和多种信息通信模态(mode)相比,严重限制了人本来所具有的通信技能。这使得人机通信效率大大低于人在自然形态中的通信效率,成为限制人机系统效率提高的一个瓶颈。近些年来多媒体信息技术的出现从技术上为在人机交互中全面采用人本来所具有的通信技能提供了可能性。但是,尽管已在许多计算机应用中引进了多媒体信息和多模态通信,人们对它们的认识仍然停留在非常粗浅的水平,人机界面的设计者往往是在缺乏理论指导的情况下凭直觉去进行设计^[1]。多媒体信息和多模态通信的复杂性使得难以用这种手工方式设计出令人满意的人机界面^[12]。本文将从多媒体信息和多模态通信的特性入手,探讨当前存在的问题以及解决问题的可能途径。

二、媒体的特性

由于人机交互作用领域的多学科交叉性质,对

于多媒体和多模态至今仍没有一致的定义。但通常认为,媒体是指用于通信的信息表示系统,如文本、语音、声音、图形、图象、录像和姿态等;而模态是指通信过程中所采用的产生和感受信息表示的手段,如语言、运动、听觉、视觉、触觉等。所谓的多媒体多模态人机交互系统(以下简称为多媒体系统)是指在人机通信中采用多种信息媒体和多种通信模态的系统^[1]。

多媒体信息在人机交互中的巨大潜力被认为^[2]主要来自于它能提高人对信息表现形式的选择和和控制能力,同时也能提高信息表现形式与人的逻辑和创造能力的结合程度,在顺序、符号信息以及并行、联想信息这二个方面扩展人的信息处理能力。多媒体信息比单一媒体信息对人具有更大的吸引力,它有利于人对信息的主动探索而不是被动接收^[13]。在动态信号与静态信号之间,人更倾向于前者。多媒体信息所提供的种类丰富的信息源恰好能满足人的这种需要。

多媒体信息有利于人机通信的重要原因还在于多媒体所带来的信息冗余性。根据香农的通信模型^[7],来自通信双方以及环境的噪声信号会对通信产生干扰。消除噪声干扰的主要方法是重复、使用别的媒体或并行使用多种媒体。多义性也是导致通信困难的一个重要原因,多种媒体的采用有助于解决这一问题。多媒体中的各种媒体有着共同的通信目

刘正捷 副教授,研究方向为人机交互作用。

标,其中的每一种媒体都会对另一种媒体所传递信息的多种可能解释产生某种限制作用,从而相互减少了多义性。

没有任何一种媒体在所有场合都是最优的。每种媒体都有其各自擅长的特定范围,在使用中必须根据具体的信息内容、通信目标、用户的偏爱以及当时的上下文选择相适应的媒体。人在问题求解过程的不同阶段对信息媒体有不同的需要^[12],相对来说,能提供具体信息的媒体适用于最初的探索阶段;能描述抽象概念的文本媒体适用于最后的分析阶段;而直观信息则介于两者之间,比较适合于综合阶段,它对设计过程尤为重要。

一般说来,文本擅长表现概念和刻画细节;图形信息则擅长表达思想的轮廓以及那些蕴含于大量数值数据内的趋向性信息。动画图形信息可用来突出某个重点(如报警信息),还适于表现静态图形所无法表现的动作信息。录像适宜于那些别的媒体所难以表现的来自真实生活的事件和情景,特别是那些需要真实感的社会文化信息,而且有助于对信息的回忆。语音能使对话信息突出,可以使人在从事某项工作的同时接收信息,它在与录像、动画结合时可以用来传递大量的信息。声音(非语音)则擅长于做系统中事件的示意、反馈信号以吸引人的注意力,还有助于激发人的想像力,音乐与人的情感有密切联系,但声音的过多使用很容易造成人的厌烦。姿态在与别的媒体结合时具有较强的信息引用能力,可以在相关信息之间建立起时间、空间以及逻辑上的联系^[6,11]。尽管如此,由于媒体资源的有限,在具体的应用场合往往要对信息内容、通信目标和表现形式(媒体)之间的关系加以调整。

信息从性质上可以分成空间性和非空间性信息二类,这里所谓的空间性信息是指有关结构、位置或运动的信息。图形媒体在表现空间性信息方面具有优越性,它可以使人在短时间内利用直觉能力接收大量的空间信息,并有利于人对空间信息的处理和记忆^[12,10]。然而也有的研究^[11]指出,媒体种类对空间信息的传递并没有明显的影响,各种媒体各有所长,文本媒体可以减少空间理解上的错误,而图形则有利于空间信息的快速接收,它们之间应取长补短,互为补充。

有的研究^[6]对自然语言(包括文本和语音)和直接操纵(包括图形和姿态)做了比较。人对图形和姿

态具有直觉性,它们与所表达的语义对象和人在语义层上的自然技能有较密切的联系,人可以以较小的认知负担对它们直接运用语义层上的知识和技能。但它们的弱点在于不擅长表达有关数量、否定和时间的信息,还缺乏指称对象和利用上下文的手段。语音和文本恰好在图形和姿态较弱的地方比较强,但它们的弱点在于存在着理解上的多义性、非直观性以及机器处理上的困难。

除了与信息内容的关系之外,媒体的选择还与用户个人的偏爱有关系。用户个人有时会由于其专业工作或特殊技能的原因而对信息的某些特定的表现形式有所偏爱。对这些个人偏爱的满足可以减少这些用户额外的认知负担,以及由此而产生的挫折感。

三、几个值得研究的问题

多媒体技术的出现为人机交互开辟了新的天地,但也带来了许多有待研究解决的问题。人机界面设计者现在面对的不是怎样(How)提供多媒体信息的问题,而是在什么情况下(When)采用什么(What)媒体及其集成以及提供相应的交互和处理手段的问题。下面将提出和探讨几个值得研究的对多媒体人机交互具有关键意义的问题。

1. 媒体的结合及其理论

多媒体界面的成功并不在于单单向用户提供丰富的媒体,而主要在于媒体的运用尤其是媒体的结合。目前的媒体结合在技术上主要是通过在一个窗口中提供多种媒体的信息片段,或是对声音、语音、录像等随时间变化的动态媒体加上同步关系来实现的。作为对于具有文本和插图信息的书本的自然扩展,这种技术主要被用于一些教育方面的应用场合,但对于如何将各种媒体加以有机的结合,仍然严重缺乏理论指导^[9]。尽管在多媒体硬件以及系统的基本控制机制方面有较大进展,但至今仍没有成熟的理论去指导人们如何在语义层上将各种媒体结合以更有效地传递信息。媒体之间可以互相支持,也会互相干扰。这里的关键问题是怎样最佳地表达某特定信息,以及怎样在多种媒体间以一种兼容和内聚的方式来分配其表达,从而在当时的上下文情况下达到最佳的通信效果。不论这一结合是在运行时由机器自动实现的,还是在设计时由人工实现的,这样的理论对于成功的人机界面设计都十分必要。目前一

些有关媒体结合的观点把问题过于简单化了。依据这些观点的媒体结合是在将各种媒体的表达彼此割裂的基础上来选择合适的媒体去表达要传递的信息。其出发点主要是利用不同的媒体去实现信息的冗余性,或是使信息能通过更有效的媒体来传递。这种观点将所传递信息的解释看成是事先确定且一成不变的。但在人的自然通信中,信息是以一种彼此相关的方式在媒体间动态分配的,而且其解释还在很大程度上依赖于当时的上下文以及通信双方的背景知识。实际上通信可以看成是由各种各样媒体的表达组成的,基于特定的通信目标,这些表达对通信双方共有的一个复杂的知识结构产生激发,从而完成信息的传递。另外,媒体间的互补性和冗余性对媒体的结合也是重要的,如果媒体间缺乏冗余,信息将失去内聚性,而过分冗余则又将导致用户集中注意某一种媒体而忽略其它。由此看来,指导多媒体通信的理论将不会是由各媒体通信理论简单合并而成。我们应当发展一种可以超脱于各媒体之上的统一的概念框架,将通信视为一种发生于特定上下文的现象,去建立和完善多媒体通信的理论。

2. 隐喻技术

在与计算机交互的过程中,人所依据的是自己所拥有的关于这种交互的概念模型(mental model)。这种概念模型的建立往往需要培训和经验。一种较好的方法是通过“模仿”(analogy)将人已有的对别的事物的知识和技能挪用过来以建立这一概念模型。这就需要在人机界面上建立起人所熟悉的某事物的隐喻(metaphor),从而使这种知识和技能的转移和复制得以顺利实现。目前在人机界面上采用较多的是来自于事务处理的“desktop”隐喻,如表格、卡片、打字机、废纸篓、信箱等。多媒体信息的出现带来了更加复杂的概念和功能,使得难以继续只用这些事务处理隐喻来帮助建立人机交互的概念模型。为此,必须发展能适应多媒体交互环境的新的隐喻技术。在这方面,“向导”^[12]和“上下文”^[13]是二个较成功的例子。隐喻技术的最终目标是要在人机交互中将机器完全隐去,这一方面要更深入地研究人在自然形态的通信中对多媒体信息的处理,从中发掘出能用于人机交互的隐喻;另一方面还要开发在概念上和物理上与之相适应的输入输出方法和设备。

人对不同信息媒体和通信模态的控制能力,还使得隐喻技术的发展可以不局限于人在自然生活中

的真实经历和感受的原原本本的重建。集成各种媒体和模态去创造新的交互方式是一条很有希望的挖掘人机交互潜力的途径。在这方面,有的研究^[15]尝试了以非语音声音当做信息直观化的主要手段,来实现满流的声觉化(audiolization)的可能性。

3. 语义

各种媒体的信息在最低层次上对于计算机来说都是二进制流。如果仅仅是做为信息的简单通道,计算机不必了解信息的语义。但若要使计算机具备媒体的选择和合成能力,则必须赋予它媒体的语义知识,从而使它能在高于原媒体表现层的层次上对不同媒体进行比较和合成。另外,为解决大量多媒体信息对存贮、传输和访问所造成的困难,也要求计算机能对多媒体信息进行语义层的处理。因此,必须研究多媒体信息语义的获取和处理。语义层是比词法、语法层更高的信息抽象层次。在获得信息语义的过程中,抽象起着十分重要的作用。这种抽象是复杂的,而且与任务相关,通常包含若干抽象层,每个抽象层都包含着与具体任务和问题域相关的模型。从接近具体感官的信息表示层到接近符号的信息表示层的抽象过程中,信息的抽象层次顺次递升,而其数据量则顺次递减。语义就是在从感官数据到符号数据的抽象过程中逐步形成的。是否可以建立类似的语义模型和抽象机制来获得多媒体信息的语义是一个值得进一步探讨的问题。

人的自然通信具有一种信息的轮廓(outline)与细节(detail)相分离的特征^[11],通常轮廓是直接由有形媒体传递的,而细节则间接地经由上下文以及背景来传递,由此实现通信的高效率。多媒体信息语义模型的实现将为建立多媒体人机通信的上下文模型,进而为在人机通信中建立轮廓与细节相分离的高效通信机制开辟一条途径。

4. 时变媒体

与文本、图形、图像这样的静态媒体不同,语音、录像、姿态等媒体是随时间而连续变化的。由于其动态性,这种时变媒体(time-varying media)比静态媒体更难以控制和处理。为此必须建立不同于静态媒体的概念和处理方法。人在自然生活中对于对话中所用到的各种时变媒体具有高度的处理技能,人在当前的对话上下文中可以灵活自如地运用插话、质询、概括、强调、引用、省略、重复以及回溯等技巧。另外,还可用手势来引用一串事件、指出

关注点以及在对象间建立某种关系。人的面部表情,特别是目光注意点也在对话中起着重要的作用。如果能在人机交互中重建这种对话隐喻,人本来具有的各种对话技能将可以用来解决时变媒体所带来的处理上的困难。

时变媒体还为在入机交互中引入戏剧成分提供了可能性。作为传递信息和增强人理解能力的有力工具,戏剧艺术在人类社会中已有几百年甚至更长的历史。如何在人机交互中利用戏剧成分来提高通信的效率是一个非常意义的研究课题。

时变媒体与静态媒体的结合,比如录像与文本的结合,也会带来新的问题。在人机交互中,录像信息一直是被分成各自包含给定帧画面的片段来处理的,这种固定的粒度与信息内容之间是不一致的。而文本信息传统上的结构则具有若干层粒度,如章、节、段、词等。这二者的结合需要对录像信息增加更加灵活的结构,从而使其访问和处理可以具有更符合其语义特点的方法,而不仅仅局限于始点、终点和帧号。

四、结束语

多媒体信息的引入为建造高效的人机界面带来了新的希望,但同时其复杂性也对人机交互作用领域提出了艰巨的研究课题。至今所开展的有关多媒体信息的研究为人们提供了许多富有启发性的观点,然而,值得指出的是,对媒体特性的认识还未达到成熟的境界,对许多问题仍缺乏共识。各个研究所得出的结论各有其特定的背景和限制,它们的普遍适用性仍有待更多研究工作的证实。在什么情况下采用什么媒体及其集成是目前人机交互作用领域所要回答的主要问题。为此,作者提出和探讨了多媒体结合、语义、隐喻技术和时变媒体处理等几个具有关键意义的研究课题。人机交互作用领域的多学科交叉特点决定了这些课题的研究必须在很大程度上借助于心理学、认知科学、语言学、通信理论、戏剧理论等多方面的理论和方法。

主要参考文献

- [1] Alty, J. L., Multimedia—What is it and how do we exploit it?, People and Computer V, D. Diaper et al. (eds.), Cambridge University Press, 1991
- [2] André, E. et al., Towards a plan-based synthesis of illustrated documents, Proc. ECAI'90
- [3] Bell, D. et al., General models of multimedia interaction, ERCIR Workshop Reports ERCIM-94-W003, 1994
- [4] Bieger, G. R. et al., Comprehending spatial and contextual information in picture-text instructions, J. Experimental Education, Vol. 56, No. 4, 1986
- [5] Blattner M. M. et al., Listening to turbulence: an example of scientific audiolization, Multimedia Interface Design, M. M. Blattner et al. (eds.), ACM Press, NY, 1992
- [6] Cohen P. R., The role of natural language in a multimodal interface, Proc. ACM UIST'92, Monterey, 1992
- [7] Fiske J., Introduction to communication studies, Routledge, NY, 1990
- [8] Laurel, B. et al., Issues in multimedia interface design: media integration and interface agents, Multimedia Interface Design, Same to [5]
- [9] Lee J., Graphics, information and multimodal HCI, Same to [3]
- [10] Levin, J. A. et al., On empirically validating functions of pictures in prose, Willows & Houghton, 1987, Vol. 1
- [11] Liu, Z. and Rada, R., A perspective of human-computer communication based on outline-detail separation, to be published
- [12] Long, J. B. et al., Conceptions of the discipline of HCI: craft, applied science and engineering, People and Computer V, A. G. Sutcliffe et al. (eds.), Same to [1], 1989
- [13] Marmolin, H., Multimedia from the perspective of psychology, Multimedia: Principles, Systems and Applications, L. Kjeldahl (ed.), Springer Verlag, 1992
- [14] Nonogaki et al., A construction of direct engagement for human interface and its prototyping, IEICE Trans. Fundamentals, Vol. E75A, No. 2, 1992
- [15] Sutcliffe, A. et al., Systematic design for task related multimedia interfaces, Information and Software Technology, Vol. 36, No. 4, 1994