Intelligent Multimedia Interfaces

Mark T. Maybury

■ *On Monday, 15 July 1991, prior to the Ninth National Conference on Artificial Intelligence (AAAI-91) in Anaheim, California, over 50 scientists and engineers attended the AAAI-91 Workshop on Intelligent Multimedia Interfaces. The purpose of the workshop was threefold: (1) bring together researchers and practitioners to report on current advances in intelligent multimedia interface systems and their underlying theories, (2) foster scientific interchange among these individuals, and (3) evaluate current efforts and make recommendations for future investigations.*

On Monday, 15 July 1991, prior to the Ninth National Conference on Artificial Intelligence (AAAI-91) in Anaheim, California, over 50 scientists and engineers attended the AAAI-91 Workshop on Intelligent Multimedia

---

Intelligent multimedia interfaces...generate output in an intelligent or knowledge-based manner.

Interfaces. The workshop was planned and coordinated by Mark Maybury (MITRE Corporation, Bedford, Massachusetts), Steve Feiner (Columbia University), Alfred Kobsa (University of Saarbruecken), and Bonnie Webber (University of Pennsylvania). The purpose of the workshop was threefold: (1) bring together researchers and practitioners to report on current advances in intelligent multimedia interface systems and their underlying theories, (2) foster scientific interchange among these individuals, and (3) evaluate current efforts and make recommendations for future investigations.

Multimedia interfaces are computer interfaces that communicate with users using multiple media (for example, language, graphics, animation, video, nonspeech audio) and sometimes multiple modes, such as written text with spoken language. *Intelligent multimedia interfaces* go beyond traditional hypermedia or hypertext environments in that they process input and generate output in an intelligent or knowledge-based manner. The multidisciplinary nature of intelligent multimedia interfaces was apparent in the workshop papers and demonstrations, which addressed a broad range of issues, spanning the disciplines of AI, computational linguistics, computer graphics, cognitive science, education and intelligent tutoring, software design, and information retrieval. Not only were the papers multidisciplinary in content, but they were also international in origin, representing research in the United States, Germany, Australia, Canada, Italy, and Japan.

The workshop was structured to address several intelligent multimedia interface issues, including multimedia design and presentation, architectures, storage and retrieval, and applications. This article describes each of these issues and points to future areas of research that were identified by workshop participants.

Multimedia Design and Presentation

The presentations in the first section of the workshop focused on methods for the automatic design of multimedia presentations. The workshop began with a videotaped demonstration of the COMET (coordinated multimedia explanation test bed) system developed at Columbia University, which automatically designs integrated textual and graphic presentations to explain the operation and maintenance of a United States Army field radio. Diane Litman (Columbia) described intended extensions to COMET for explicit representations of temporal relations underlying domain actions and plans. Kathy McKeown (Columbia) discussed issues in generating speech (temporal media) about temporal events, which included planning content in a given period of time as well as using temporal markers, such as the adverbials "before" or "while," to convey information about temporal relations of underlying domain events. These markers (together with tense and aspect) are particularly useful when the presentation order is not isomorphic to the underlying task order. Finally, Steve Feiner (Columbia) discussed methods for the automatic design and coordination of three-dimensional graphics as well as the incorporation of animation into COMET.

Elisabeth Andre (German Research Center for AI) detailed the interdependence of the generation of language and graphics, focusing on generating unambiguous linguistic and visual references to objects. Winfried Graf (German Research Center) detailed a constraint-based approach to document layout used in the system WIP (knowledge-based information presentation [trans. from German]), specifying design principles such as gridding, alignment, and symmetry as well as a technique for propagating obligatory, optional, and default constraints to position individual document fragments on a design grid. John Burger (MITRE) described an intelligent multimedia interface (AIMI) that can engage a user in a multimedia dialogue, for example, responding to a natural language query by automatically designing business-like graphics, which the user can then interact with or refer to. Ralph Marshall (MITRE) commented on AIMI's ability to choose alternative media to express information from an underlying KL-ONE–like knowledge base, for example, using nonspeech audio to convey the speed, stage, or duration of an otherwise invisible process.

Joe Marks (Harvard University) proposed notions of syntactic, semantic, and pragmatic coherence to help ensure consistency in the design of node and arc diagrams. He discussed extensions to his system, ANDD (automated network diagram designer), to include a design history that would

ensure design consistency in sequences of network diagrams. Angel Puerta (Stanford University) argued that interface design excellence was an incorrect goal; instead, we should focus on *self-adaptive interfaces*, namely, those that learn to be effective based on interactions with users performing specific tasks in particular domains.

Yigal Arens (USC Information Sciences Institute [USC/ISI]) discussed work he performed with Eduard Hovy (USC/ISI) and Mira Vossers (USC/ISI) that aims at categorizing the range of knowledge required to reason about multimedia presentations, including knowledge about the kind of information to display, the goal of the producer, the characteristics of the reader, and the nature of the media (for example, text versus graphics). Arens analyzed a multimedia exposition from an automobile manual of how to adjust the front seat of the Honda Accord. He detailed how distinctions concerning the content, form, and purpose of manual components might be captured in a systemic network.

Jurgen Krause (University of Regensburg) argued for the need for empirical validation of multimedia interfaces to determine when, how, and why certain media should be chosen for communication. Using carefully designed "Wizard of Oz" experiments (experiment of interaction between a human and a simulated computer; analogous to the interaction between Dorothy and the simulated wizard in the land of Oz) comparing natural language with graphic interfaces, Krause is currently investigating the hypothesis that different classes of computer talk exist for different kinds of interfaces.

Multimedia Design and Presentation Issues

Steve Feiner (Columbia) highlighted a number of issues that were raised during the initial sessions of the workshop. For example, it was evident from the presentations and discussions that multimedia design involves a number of complex issues, including temporal coordination of multiple media, the relationship of textual and graphic generation, automatic design of graphics, and modality selection (for example, realizing language as text or speech). Another issue concerned the degree of automa-

Another issue concerned the degree of automation versus mixed initiative.

tion versus mixed initiative. For example, the WIP system focuses on the design of stand-alone instructions, but AIMI allows interaction through a presentation's mouse-sensitive components, and COMET allows the user to control camera positions. Other issues concern whether systems save the history or structure of a presentation and if and how animations are connected to representations of abstract knowledge. It was pointed out that the need for deep knowledge of designed graphics depends on the intended use of the multimedia presentation (for example, teaching versus manual generation) and the environment in which it is used (for example, interactive, static).

Several common problems were identified with respect to media coordination. These problems include the need for presentation balance, mutual reference and the interaction between text and graphics, and the relationship between the characteristics of the information to be presented and the devices available for presentation. It was noted that WIP actually exploits two feedback loops, one after presentation design and one after realization, to help resolve intermedia and intramedia synthesis problems. Related to the need to dynamically plan presentations is the choice between plan reuse, refinement, and replanning after a failed presentation. In addition, when multiple choices among presentations are possible, a common problem raised was the need for goodness metrics that, for example, measure the consistency and coherency of multimedia presentations.

Architectures

A prerequisite to multimedia presentation design is architectures for multimedia information processing. Several presentations described frameworks for multimedia interfaces and multimedia information systems. Matthew Cornell, Daniel Suthers,

and Beverly Woolf (University of Massachusetts at Amherst) showed a videotaped demonstration of their multimedia research into an interactive environment that aims at developing a framework for sharing information and viewing it from a number of different perspectives to support a philosophy of learning by research. They demonstrated how a student could query their system; get a response; and then edit this response or place it into various viewers that would, for example, display the information as text, graphics, or an animation.

Tetsuro Muranaga (Toshiba R&D Center, Japan) discussed an integrated multimedia and software architecture, designed together with Rajiv Trehan (Toshiba), that was based on an object-oriented database and concurrent object-oriented C. Thomas Hemmann (German Institute for Applied Information Technology) described a media assistant application.

Storage and Retrieval

The need to efficiently store and retrieve multimedia information was also discussed, encompassing issues of multimedia representation, indexing, and integration. Marc Davis (Media Laboratory, Massachusetts Institute of Technology) described the use of semantic logging of videotaped information using intelligent icons that were related to one another in an attempt to support a "director's workshop." The goal of this work was to provide more semantically oriented video-logging tools to support subsequent retrieval and resequencing of video for automatic presentation.

Guy Boy (NASA Ames Research Center) then described an intelligent browsing system he developed, together with Cecile Paris (USC/ISI), that addresses the problem of browsing through large technical documentation by using a device called a "contextual link" to make hypertext links context sensitive. Boy demonstrated the system CID (computer integrated documentation), which utilizes online feedback from the user to reinforce existing contextual links if a user's request succeeds or generate new ones if it fails.

Applications

A final section of presentations addressed a range of multimedia

applications. James Ragusa (University of Central Florida) described his practical experience building multimedia applications for the National Aeronautics and Space Administration that address, among other things, the need to classify and retrieve information from 4 million still images of the space shuttle prelaunch processing cycle. Andrea Bonarini (Milan Polytechnic) described research to support a multimedia interface for a driver in an automobile. He discussed some design criteria for effective input and output tools and the need to model a number of properties of the driver (for example, his/her psychological state, attitudes, preferences, knowledge, and goals) and described an architecture for a multimedia car-driver interface.

Mark Guzdial (University of Michigan) then described and presented a videotape of work he engaged in with Elliot Soloway (University of Michigan) to develop an educational multimedia composition environment. This system has been used in a number of subject areas, including physics, composition, and journalism, at Community High School in Ann Arbor. Guzdial and Soloway are currently exploring the extension of their system, MEDIA TEXT, to capture design information and connections between text and media (for example, images, video, animations). Finally, Bob Barbour (University of Waikato, New Zealand) presented a videotape that argued that different individuals prefer different media for information presentation.

Future Directions

The workshop concluded with a discussion led by Alfred Kobsa (University of Saarbruecken) that examined a number of areas for future work in intelligent multimedia interfaces. One area concerns when, how, and why media are chosen to convey different types of information. Several participants took a practical approach to this problem, building systems that are based on reverse engineering of naturally occurring presentations; however, Angel Puerta argued that media selection should be a machine-learned activity based on interaction with users, and Juergen Krause argued that it requires empirical validation through observation of man-machine interactions. Related to this focus on empiricism is the need to provide statistical evidence that the

---

Another issue concerns… the need to capture the syntax, semantics, and pragmatics of graphics.

---

additional machinery required to design and render more complex multimedia presentations is warranted by some pedagogic benefits, increases in efficiency, or increases in the effectiveness of accomplishing some task.

Another issue concerns the well-formedness of graphics, encompassing the need to capture the syntax, semantics, and pragmatics of graphics. First, research in this vein is already being carried out. Syntactic well-formedness criteria presented at the workshop included rules for design consistency (Marks), perceptual grouping (WIP), and contrast enhancement (Arens, Feiner). Pragmatic criteria included restrictions for reference (Andre) and persistence rules (WIP).

Still to be resolved is the storage and retrieval of multimedia information. Kobsa pointed out that we should learn from the information and library sciences, which have concentrated for decades on the classification, abstraction, storage, and retrieval of (primarily text) documents, and investigate analogies between document processing and multimedia information processing.

Finally, Kobsa pointed out that the capabilities of intelligent multimedia systems presented at the workshop go beyond hypermedia to include the ability to interpret (possibly multimedia) questions and automatically design multimedia answers (for example, WIP, COMET, AIMI, andd), deal with following up questions and make backward references (for example, AIMI), postedit presentations (for example, COMET), and intelligently index and retrieve multimedia information (for example, CID). Other areas that require further research include the incorporation of dialogue (for example, context and turn taking) into multimedia interfaces, more complex pedagogic models

(such as Cornell, Suthers, and Woolf's perspective of learning as research), and a capability to provide diagnosis and advice as a user designs a presentation (for example, using a CAD tool). A final suggestion concerned tailoring multimedia presentations to individual user's knowledge, abilities, preferences, goals, and plans. The research results on reader adaptation in technical documentation, user modeling in interactive computer systems, and intelligent tutoring systems might also be relevant to this endeavor.

Summary

It was clear from the lively workshop discussions that there were many unresolved issues shared by the participants. The range and depth of the papers submitted to the workshop reflected the magnitude of the problems surrounding intelligent multimedia interfaces but also the current interdisciplinary investigations into many of the key issues. Although research in this area is still in its formative stages, the organizers hope that the workshop will help foster the scientific interchange necessary to solve many of the remaining fundamental problems.

To obtain the full addresses of any of the participants, please contact Mark Maybury, Mail Stop K329, Artificial Intelligence Center, MITRE Corporation, Bedford, MA 01730. Telephone: (617) 271-7230. Fax: (617) 271-2352.

Mark T. Maybury is group leader of speech and natural language technology at MITRE Corporation in Bedford, Massachusetts. Maybury's Ph.D. thesis at Cambridge University (1991) was entitled "Planning Multisentential English Text Using Communicative Acts." His interests include natural language processing, planning, and knowledge-based simulation. His current research focuses on the development of intelligent multimedia interfaces, particularly the representation and use of communicative acts for designing multimedia documents and interpreting and generating multimedia dialogue.