MULTIMODAL COMPUTING ENVIRONMENTS: THEORY AND APPLICATIONS

(NSF Contract #IRI 9213823)

Meera M. Blattner*
Ephraim P. Glinert†

Program Area: Adaptive Human Interfaces

Keywords and Phrases: multimodal display spaces; multimodal display algorithms; multimodal UIMS; metawidgets, pen and voice; pen systems; earcons; audio management.

1 Area Background

Multimedia systems try to take advantage of human senses to facilitate our communication with one another or with the computer. The desire to achieve natural and expressive interaction has given rise to many input devices that transform human expression, such as gaze, voice, gesture, to bits and bytes. A multimodal interface exploits different user senses, such as touch, hearing, and seeing.

The greatest innovations in hardware and software development today stem from advances in our understanding of the ways people use computers. The changing focus of interface development has shifted research from perceptual and motor issues to cognitive issues centered on learning and using systems. Multimedia interface developers face the additional problems of integrating media and designing models for time–varying systems.

2 Project Summary

2.1 Introduction

The proliferation in recent years of powerful and affordable interface technology has led to a mushrooming interest in so–called multimodal or multimedia environments. Exploration of this new and powerful type of environment is still in its infancy. In particular, attempts to rigidly allocate certain output to the visual channel or to the audio channel will not accommodate users' cognitive requirements. Some users need to accommodate disabilities. Conditions in the workplace often impose limitations on what kinds of output are acceptable. Users can find it difficult to understand certain information when it is presented in the "wrong" modality.

Viewing the visual channel alone as one extreme of a continuous spectrum and the audio channel alone as the other, users should be able to control the proportion of information they are receiving via each, and perhaps its nature as well. Indeed, real–world objects (which will be important in virtual realities) can and do undergo metamorphoses. Some can normally be heard but not seen (e.g., the engine in our car as we drive), while others can be seen but not heard (e.g., the telephone perched on our desk). However, when we lift the car's hood we can see the engine; when the phone rings we can hear it. So, as the work session progresses, it must be possible for both system and user to revise the initial choice of display modality for given information, as conditions may dictate.

Observations such as these led us recently to propose a new class of objects in the interface called *metawidgets*. These are abstractions of the widgets with which we are familiar, in that they are clusters of alternative representations for some "information packet" (e.g., a menu, a graph, an image, an icon, etc) along with built–in methods for selecting among

*First author's contact information: University of California at Davis & Lawrence Livermore National Laboratory, Livermore CA 94550; blattner@llnl.gov.

†Second author's contact information: Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180; glinert@cs.rpi.edu.

these representations. Metawidgets naturally define an object oriented hierarchy, according to the type of information they are designed to convey to the user. The ways in which information packets can manifest themselves in a given environment comprise a second object oriented hierarchy, in which there is potential multiple inheritance of properties relating to each of the sensory modalities that the environment is capable of stimulating.

The metawidget team of six members from the University of California, Davis (UCD), and Rensselaer Polytechnic Institute (RPI). The uniform hardware platform for all work described above consists of Pentium–based IBM PC's or compatibles. Team members are in email communication; six of them (Meera M. Blattner, UCD; Ephraim P. Glinert, RPI; Richard L. Kline, RPI; Andre D. Milota, UCD; Albert L. Papp III, UCD; and G. Bowden Wise, RPI) met at CHI '95 in Denver where extensive discussions were held. Kline, Milota, Papp, and Wise are doctoral students at their respective institutions. Another team member, Joaquim A.P. Jorge, RPI, recently completed his dissertation.

2.2 The Architecture of Metawidgets

Wise has defined and iteratively refined the fundamental architecture of metawidgets based systems. He has also written detailed specifications for the various core software components of such systems, including the resource consumption monitor, the presentation manager, the multimodal presenter libraries, and the user preferences files.

- The resource monitor is responsible for determining the contribution to the total cognitive load imposed by each application running in the environment. The tricky part is to measure those attributes which are meaningful in terms of people, yet which may not coincide with events easily detectible by or values readily available to the computer.
- The presentation manager's task is to assure that the information on display is distributed appropriately among the sensory modalities so that the user does not become overloaded. In some cases this will necessitate the transformation of information to an alternative modality; the system is expected to handle such chores in a transparent manner.

Both of these tools will run in the background and coexist with user metawidget applications. Each metawidget contains a set of representations (usually in different modalities) and methods to select among them.

2.3 Metawidgets for Pen and Voice

In parallel with this effort, Jorge extended his work on the use of gestures and pen based input, and was able to obtain significant results (both in terms of concepts and algorithms) relating to the use of gestural input. In particular, he has developed a fast incremental parsing algorithm for a broad class of visual "languages" which contain no overlapping elements. The algorithm is based on the concept of adjacency grammars, where adjacencies are defined so as to encompass both spatial and logical constraints. Measurably efficient online parsing of diagram–like input, with observed linear run times for large (visual) sentences, is achieved. Other powerful recognition algorithms are based on ideas derived from fuzzy logic, which allows Jorge to quantify degrees of ambiguity and imprecision in visual sentences in a more meaningful way than is possible using crisp formulations.

In his SCRIBBLE system, Milota is investigating the combined and complementary use of speech and gestures (via a pen) for input. The goal is to understand the mechanisms that, at least for certain application classes, allow us to achieve results superior to those attainable using either input modality alone, yet adapting each modality to the context of its environment. SCRIBBLE is natural language with a limited vocabulary combined with gesture to communicate with the computer. Each command or query has several different forms. Whenever possible, one of these is a purely gestural version and another a pure voice version. Various combinations of the two modalities are also provided where appropriate.

2.4 Metawigets and Audio

Papp and Kline are both exploring aspects of aural input.

- Papp is seeking to understand how we can concurrently present multiple information items aurally, so that those which are "more important" receive suitable prominence over those which are "less important" without obscuring them altogether. He plans to use such knowledge as a foundation for designing new and effective aural toolkits for metawidgets

- Kline is exploring the potential of real time analysis of nonspeech audio input from the environment, coupled with "secondary" speech characteristics such as inflection, pauses and monotony, to allow the computer to experience its world by listening to it in much the same way humans do.

To better explore both informational encoding in sound and user perception of concurrent audio presentations, Papp has implemented a prototype (and as yet unnamed) system which allows the same object to be represented in various auditory structures including speech, earcons, and auditory icons (real world sounds). When a client application requests that an object be displayed, the available system chooses one of the representations but may then possibly alter it before presentation to the user. Decisions about when (in time) and in which form an auditory object is to be displayed are based upon consideration of parameters such as latency, priority, and user preferences.

Although intended for multimedia systems designed for the general user community, we envisage that metawidgets will play an especially central role in future assistive technology systems. Thus, Wise presented an overview of his work this past summer at the 18th Annual Conference of the Rehabilitation Engineering Society of North America (RESNA) in Vancouver. Furthermore, Glinert (together with Shi–Kuo Chang of the Computer Science Department in the University of Pittsburgh) is organizing a workshop on the topic "Adaptive Multimedia Technologies for People with Disabilities" in San Francisco in conjunction with ACM's MULTIMEDIA'95 conference.

3 Project References

Publications generated during the past year include the following:

1. Ephraim P. Glinert (moderator), Meera M. Blattner, Shi–Kuo Chang and David J. Kurlander: Visual Languages and Programming in the Year 2004 (panel position paper). In *Proc. 1994 IEEE Symposium on Visual Languages (VL'94)*, St. Louis, October 4–7, pp. 162–166.
2. Meera M. Blattner, Albert L. Papp III and Ephraim P. Glinert: Sonic Enhancement of Two Dimensional Graphic Displays. In "*Auditory Display: Sonification, Audification and Auditory Interfaces*" (Gregory Kramer, Editor), pp. 447–470. Santa Fe Institute Studies in the Sciences of Complexity, Volume XVIII. Addison Wesley Publishing Co., Reading (Mass.), 1994.
3. Meera M. Blattner: In Our Image: Interface Design in the 90's. *IEEE Multimedia* **1**(1):15–26, 1994.
4. Richard L. Kline and Ephraim P. Glinert: Improving GUI Accessibility for People with Low Vision. In *Proc. 12th ACM SIGCHI Conf. on Human Factors in Computing Systems (CHI'95)*, Denver, May 6–11, pp. 114–121.
5. Albert L. Papp III and Meera M. Blattner: A Centralized Audio Presentation Manager. In *Proc. 2nd International Conf. on Auditory Display (ICAD'94)*, pp. 231–237. Santa Fe Institute Studies in the Sciences of Complexity. Santa Fe Institute, 1995.
6. G. Bowden Wise and Ephraim P. Glinert: Metawidgets for Multimodal Applications. In *Proc. 18th Conf. of the Rehabilitation Engineering Society of North America (RESNA'95)*, Vancouver (Canada), June 9–14, 1995, pp. 455–457.
7. Joaquim A.P. Jorge and Ephraim P. Glinert: Online Parsing of Visual Languages using Adjacency Grammars. In *Proc. 1995 IEEE Computer Society Symposium on Visual Languages (VL'95)*, Darmstadt, September 5–9, 1995, pp. 250–257.
8. Meera M. Blattner: Multimedia Interfaces: Designing for Diversity. To appear in "*Multimedia Systems and Applications*" (Borko Furht, Editor). Kluwer Academic Publishers, Boca Raton, 1995.
9. Meera M. Blattner and Andre D. Milota: Multimodal Interfaces using Pen and Voice Input. In *Proc. IEEE International Conf. on Systems, Man and Cybernetics*, Vancouver, October 1995.

4 Related Program Areas and Potential Projects

Any one of the six program areas fit well with the Metawidgets project. Part of the work mentioned is with speech communication, particularly dialogue models and how they are integrated with other modalities. Nonspeech audio has been central to the work Blattner and her students are doing for the last ten years. Glinert is a leader in the field of computers for those with disabilities and several current projects are in this area of research. The metawidgets team makes usability studies of the prototyped projects, in particular, Blattner has several students (not mentioned above) that are using contextual design and ethnographic methods for interviewing users. These and other techniques used in human factors analysis can be used in studying metawidgets as well.