Virtual Spaces Revive Real World Interaction

Luc JULIA, **Jehan BING** and **Adam CHEYER**
Computer Human Interaction Center (CHIC!) - SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025 – USA
{luc.julia,jehan.bing,adam.cheyer}@sri.com
http://www.chic.sri.com

Abstract

As virtual spaces become more realistic, researchers are experimenting with new perspectives for interactions with such environments. Based on several prototypes that explore augmented and virtual reality as well as dialogs with lifelike computer characters, we discuss in this paper future directions for virtual environment interfaces that look back to the "good old ways" of working in the real world: talking, gesturing, moving, drawing, and so forth.

Introduction

In today's computing world, almost all users interact with their computers through the same set of input/output devices – a keyboard, mouse, monitor – and user interface metaphors – a "desktop" on which "folders" and "files" are organized, and "windows" that provide manageable views into larger spaces (documents). However, as Moore's Law makes exponentially greater computational power available to the masses, we expect that new interfaces and paradigms will emerge. What will these be like? It is our hypothesis that for advanced new interfaces to become successful and ubiquitous, they must be made simple and familiar in nature, echoing experiences in real life.

Augmented Pen and Paper Metaphor

Our first work with extended input peripherals and alternative interface metaphors focused on adapting a user's interaction with a pen and piece of paper to the electronic realm. In the TAPAGE/DERAPAGE applications [Figure 1, Left], a user would conceptualize a complex nested table or flowchart, draw a rough freehand sketch of the concept, then engage in an interactive dialog with the system until the desired product was realized [6]. Interactions consist of natural combinations of both pen and speech input – a user can cross out an undesirable line, draw in new additions, and reposition lines or objects using commands such as "put this over here." In these applications, we tried to capture the nature of a pen/paper experience, while enhancing the paper's role to become a partner in the process, capable of following high-level instruction and taking an active role in the construction of the document.

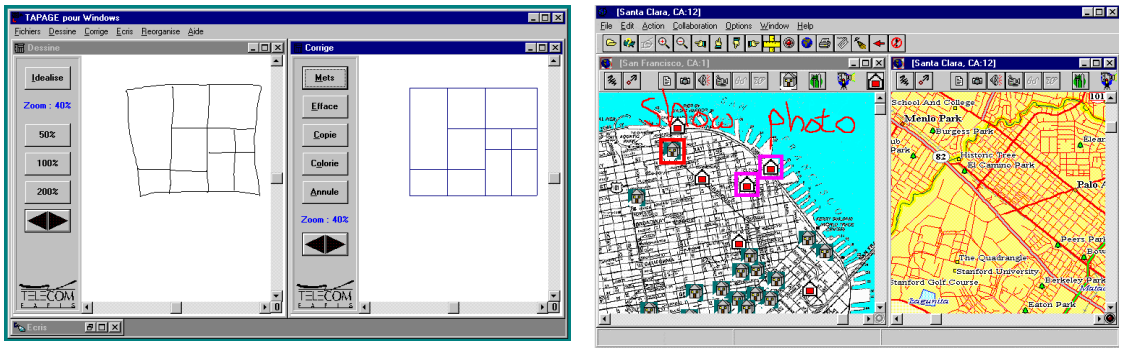

Figure 1. TAPAGE and MMap: Interactive Paper and Maps using Pen and Voice

A second project focused on applying the metaphor of "smart paper" to the domain of maps, where the goal is to manipulate and reason about information of a geographic nature [Figure 1, Right]. Inspired by a simulation experiment described in [7], we developed a working prototype system of a travel planning application, where users could draw, write, and speak to the map to call up information about hotels, restaurants, and tourist sites [4]. A typical utterance might be: *"Find all French restaurants within a mile of this hotel" + <draw arrow towards a hotel>*.

The research challenges in constructing such a system are in how to develop a multimodal engine capable of blending incoming modalities in a synergistic fashion, and able to resolve the numerous ambiguities that arise at many levels of processing. One problem of particular interest was that of reference resolution (anaphora). For example, given the utterance "Show photo of *the* hotel", several distinct computational processes may compete to provide information: a natural language agent may volunteer the last hotel talked about, the map process might indicate that the user is looking at only one hotel, and a few seconds later, a gesture recognition process might determine that a user has drawn an arrow or circled a hotel. To better understand these factors, we constructed a set of user experiments based on a novel variant of the Wizard of Oz (WOZ) simulation methodology called the WOZZOW[1] technique. These experiments are run in such a way that we can gather data from a user population, analyze the data, and directly adapt our working prototype based on the results, quantifying how much findings actually improve the system [5].

3D Paper Metaphor?

Through the previous experiments and constructed systems, we were able to develop some sense of how a "smart paper" metaphor could be brought to 2D tasks. However, with 3D becoming more prominent in user interfaces [2], we were thus curious whether the same input techniques (i.e., drawing, writing, speaking) would be effective for 3D situations.

[1] WOZZOW is a palindrome representing a single experiment with two halves, the WOZ side, which is a standard Wizard of OZ simulation experiment, and the ZOW side, where an expert user receives queries from our WOZ subject, and using a fully automated version of the simulation, tries to produce the desired effect as fast as possible, to make the WOZ subject believe he is using a real system.

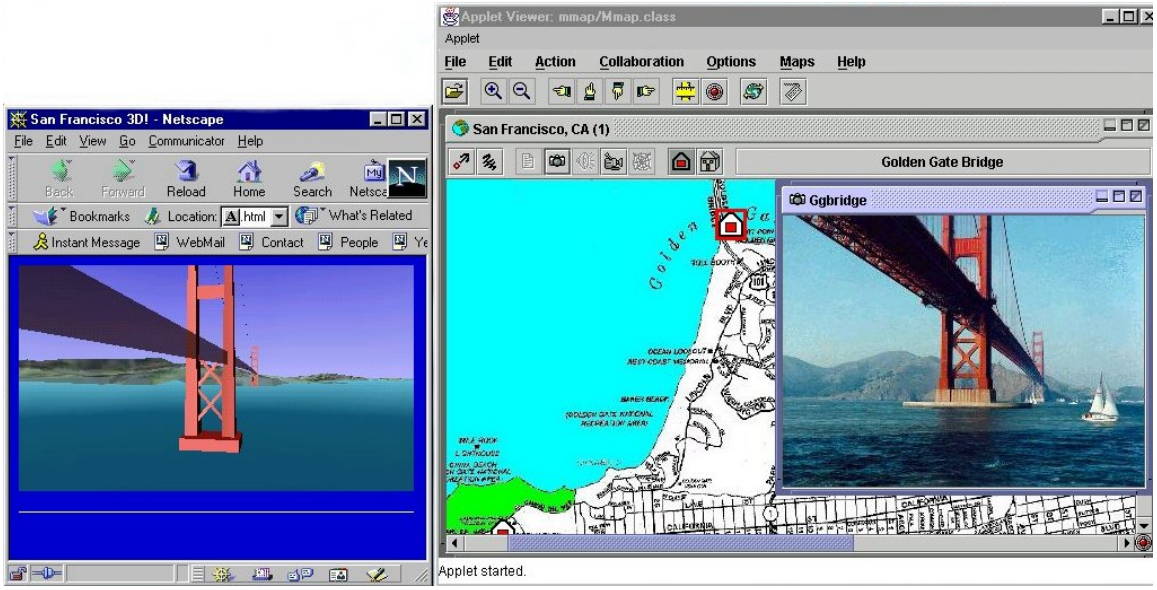

Figure 2: Multimodal interactions in synchronized 2D & 3D maps

To create an environment in which to pursue this investigation, we began by augmenting our 2D map by a 3D virtual reality (VR) view of the world [Figure 2]. A user can choose to interact with this system using pen and voice in either a 2D window (map – bird's eye view) or a 3D window, and the two are kept synchronized, with viewports and object information icons updated simultaneously in both.

Although many commands remain primarily the same in both 2D and 3D worlds (e.g., "Bring me to the Hilton"), it is unclear how to best interpret both pen gestures and speech utterances for 3D. For instance, does an arrow to the left indicate the user wants to turn towards the left, keeping the same position, or rather pan her position towards the left, keeping the same orientation? What does the spoken reference "up" mean in the context of complex 3D terrain. Although clearly a 2D paper metaphor doesn't transparently map onto a 3D environment, we have begin conducting more detailed experiments focusing on pen-voice interactions for 3D models, specifically looking at:

- Deictic and gestural reference to features of the terrain: How do people refer to and distinguish between features of a terrain model with words and gesture?

- Discourse structure: How does the structure of the interaction enable more economical communication, and how can a computer system utilize this structure in interpreting spoken and gestural input? How is the discourse structured by the structure of the terrain model and of the task or operation being executed in the terrain?

- Spatial language: How does language carve up space, and what is its relation to more geometric representations of space used in terrain models?

In addition to these experiments, we have been exploring speech recognition in conjunction with other mechanisms for navigation in a 3D virtual world. An initial prototype [Figure 3] explores the use of speech to allow higher-level expression of navigational intent for piloting a virtual vehicle (e.g., "Follow this shark."). We believe that this form of interaction will have an impact in the 3D gaming arena, and are planning on investigating these possibilities more closely.

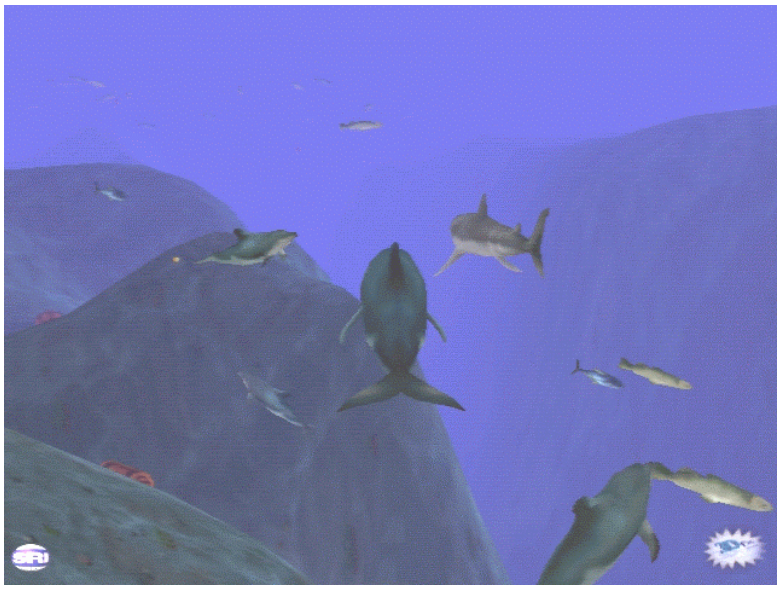

Figure 3: Immersion in Virtual World

Augmenting the Real World with a Virtual World

Although pen and voice input seems potentially promising devices for interacting with 3D environments, we are looking for solutions that provide less intrusive and even more natural interactions. Sensors are now becoming available that allow computer systems to monitor a user's position, orientation, actions, and views, and construct a model of the user's experience. Access to such a model will enable computer programs to proactively and continually look to enhance the user's real-world perceptions, without specific intervention from the user. This concept is popularly known as "augmented reality" (AR).

To enable exploration of the augmented reality paradigm, we have been constructing an AR application framework, called the Multimodal Augmented Tutoring Environment (MATE). In this framework, multiple processes for providing sensor readings, modality recognition, fusion strategies, viewer displays, and information sources can be quickly be integrated into a single flexible application. Our first AR prototype "Travel MATE" [Figure 4] makes use of many of the technologies developed in our 2D and 3D tourist applications, but adds GPS and a compass sensors. As a user walks or drives around San Francisco, a small laptop computer or PDA simultaneously displays a 3D model of what they are seeing in the real world, automatically updated

based on the user's position and orientation [8]. If the user wants to know what a particular building in the distance is, she can look at the display where objects in view are labeled. More detailed multimedia information about these objects can be retrieved on request.

The goal of the Travel Mate application is to provide useful contextual information to the user in an unobtrusive way. We are also working on an "Office MATE" prototype to investigate how AR could enhance the workplace.

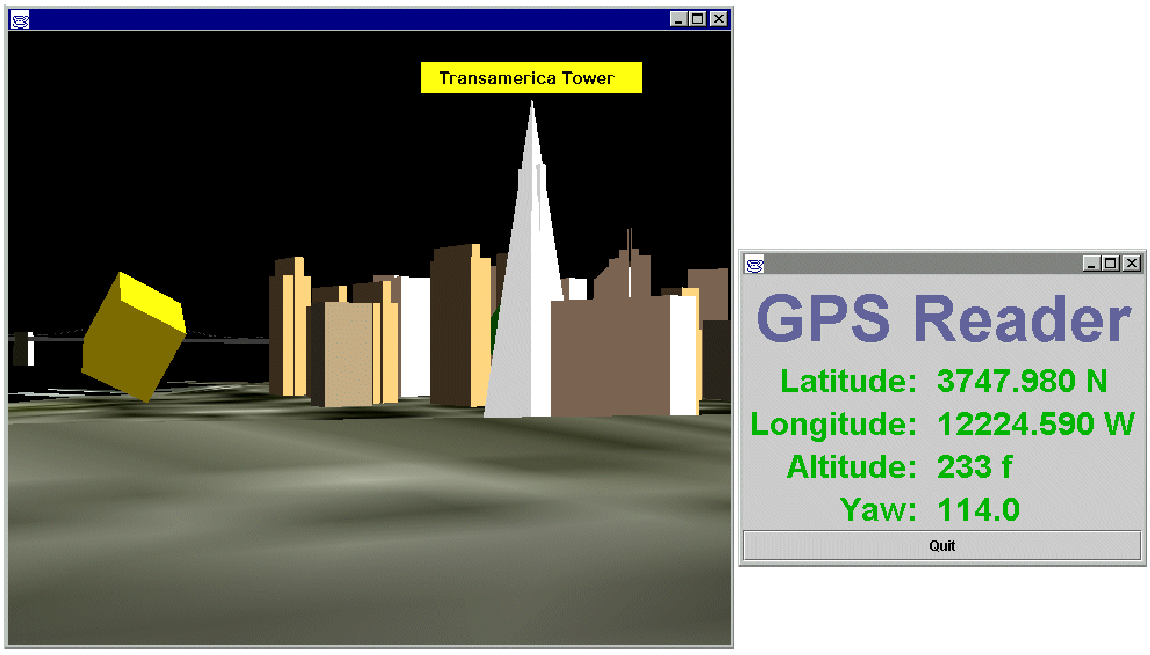

Figure 4: Travel MATE, easy and natural access to touristic information

Interacting with a Social Computer

Many interchanges between people involve lively, two-way conversations that make use of spoken dialog. The communication styles between humans and today's computer programs however are often much more restricted, with the user directing and the computer passively following orders. We feel that future user interfaces must explore a larger space of interactions with more varied ranges of participation from both sides. In our Travel MATE prototype, we saw an attempt to have a more proactive provision of information from the computer. In our InfoWiz Kiosk application, we look at other interaction styles between human and machine.

The InfoWiz project is centered around the idea of putting an interactive kiosk into the lobby of SRI [3]. People who have a few minutes to spend will be able to learn something about the institute, enjoy themselves, and hopefully walk away with a good feeling of having seen something interesting and unusual.

As users approach the kiosk, they are presented with a web browser containing information about SRI, and an animated cartoon character known as the InfoWiz [Figure 5]. Instead of using a touch screen or mouse to navigate through the information, all interactions with the kiosk occur through spoken requests issued into a telephone (a real-world, familiar interface). As users browse the InfoSpace, the InfoWiz Wizard can observe their actions, provide supplementary information, answer questions, take users on guided tours, and otherwise engage the user in a dialog about what they are seeing and about SRI.

Research issues in constructing such a system involve: how to codify, populate and maintain the InfoWiz's knowledge about the target web pages; what types of dialog structures will emerge in such a domain; what social cues must the InfoWiz follow when interacting with a user and how do they change across contexts and users; how to maintain the illusion of intelligence given imperfect recognition technologies and inadequate knowledge. These topics must be explored given the challenges of a domain where users will be from a very diverse population (many people visit SRI) and where there is no time to train the users about the system's capabilities – total interaction time with the system is expected to be a few minutes. In addition to our own approach, several good solutions can be found in research such as [1].

Figure 5: InfoWiz, SRI's interactive kiosk

Conclusion and Future Directions

The metaphors we use today to interact with computers were developed primarily in the 1960's and 1970's by researchers from SRI and Xerox. As computers, sensors, bandwidth, display capabilities, and software techniques continue to improve at incredible rate, providing computational power only dreamed of during the 60's and 70's, opportunities are emerging to transform the paradigms used in human-computer interaction. However, we feel it important to reemphasize that future interfaces can learn

a lesson from the longevity of keyboards and desktops – interfaces will be more readily adopted by the population of users if they are simple, natural, intuitive, and familiar.

In this paper, we have discussed some of our research efforts directed at attaining this goal, focusing on techniques for applying the metaphors of "smart paper" to 2D and 3D environments, creating multimodal interfaces for virtual and augmented reality, investigating the use of mixed-initiative spoken language dialog with it's implications for social roles between humans and machine. Although much progress has been made in our group and elsewhere, creating "simple" interfaces is still not a simple problem, and much research remains.

As a closing remark, we would like to comment that the speed with which new technologies are emerging is faster now than at any time in our history. It's interesting to note that many of the familiar objects on which today's interface paradigms are based have already been replaced: for instance, although keyboards are everywhere, it is no longer easy to find a typewriter. Will electronic pens and paper eventually replace their real versions? If the world is changing so fast that nothing has time to become familiar before it is replaced by something else, how will our society be able to deal with the pace?

References

[1] E. Andre, T. Rist, and J. Muller. Guiding the User Through Dynamically Generated Hypermedia Presentations with a Life-Like Character. In *Proceedings of International Conference on Intelligent User Interfaces (IUI-98)*.

[2] W. Ark, C. Dryer, T. Selker and S. Zhai. Representation Matters: The Effect of 3D Objects and a Spatial Metaphor in a Graphical User Interface. *CHI'98*.

[3] A. Cheyer and L. Julia. InfoWiz: An Animated Voice Interactive Information System. In *Proceedings Agents'99, Workshop on Conversational Agents and Natural Language*.

[4] A. Cheyer and L. Julia. Multimodal Maps: An Agent-based Approach. In book *Multimodal Human-Computer Communication, Lecture Notes in Artificial Intelligence #1374*, Springer.

[5] A. Cheyer, L. Julia and J.C. Martin. A Unified Framework for Constructing Multimodal Experiments and Applications. *CMC'98*.

[6] L. Julia and C. Faure. Pattern Recognition and Beautification for a Pen Based Interface. *ICDAR'95*.

[7] S. Oviatt. Multimodal interfaces for dynamic interactive maps. *CHI'96*.

[8] http://www.chic.sri.com/projects/MATE.html