Facilitating Navigation and Information Access with CANOES: Collaborative Agents and Natural Operations in Enhanced Spaces

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

Keywords: Multimedia, Multimodal, Proactive, Augmented Reality

Introduction

The space around us contains information of different types: local, contextual and global. By enhancing in an unobtrusive way a person's interaction with the space they live in, we aim to supplement this information both in a proactive way according to a user's interests and tasks, and through "natural" requests. In the CANOES project, we are developing several advanced interactive prototypes to study information access and navigation using natural interactions in spaces augmented by collaborative agents. The results of these studies will feed back to enhance the performance of individual technologies integrated to produce the prototype, thus improving the global system. This process is continually iterated to produce a bootstrap effect.

Collaborative Agents

Over the past few years, SRI has developed a general-purpose framework, called the Open Agent Architecture (OAA), for building distributed applications from dynamic communities of heterogeneous software agents. OAA is structured so as to minimize the effort involved both in creating new agents and in "wrapping" legacy applications; to encourage the reuse of existing agents; and to allow for dynamism and flexibility in the makeup of agent communities. Distinguishing features of the OAA include facilitator-based delegation of complex goals, temporal control and data management facilities that may be used consistently both within and between agents, and built-in support for including the user as a privileged member of the agent community.

Natural Interfaces

Metaphors

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 can 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 is realized [4]. 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 part in the construction of the document.

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 [6], we developed MMap, 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 [1]. A set of collaborative agents helps the user to find the right

information through a reactive, multimedia, interface. A typical utterance might be: *"Find all French restaurants within a mile of this hotel" + <draw arrow towards a hotel>*.

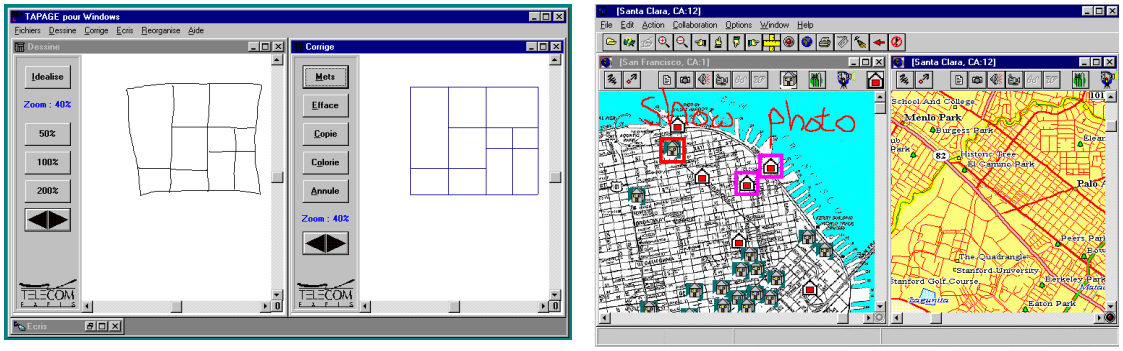

Figure 1. TAPAGE and MMap: Interactive Paper and Maps using Pen and Voice

We further adapted the multimodal map to other domains, including tasking a set of semi-autonomous robots [3] using a "coach metaphor" [5].

Synergies and Ambiguities

The research challenges in constructing such systems 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.

Evaluation

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 technique. 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 the real working prototype, tries to produce the desired effect as fast as possible, to make the WOZ subject believe he is using a real system. 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 [2]. This results in a simple bootstrap loop.

Augmented Reality for Smart Environments

Although pen and voice input seems potentially promising devices for interacting with 2D environments, we are looking for solutions that provide less intrusive and even more natural interactions in our 3D space. 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 quickly be integrated into a single flexible application. Our first AR prototype "TravelMATE" [Figure 2] makes use of many of the technologies developed in our 2D tourist applications, but adds GPS and a compass sensors for the 3D navigation. 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]. In the current

version, 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. When a good quality, see through, heads up display will be available, the alignment of the virtual image will naturally augment the real world. More detailed multimedia information about these objects, contained in the virtual world, can be retrieved upon request. The goal of the TravelMate application is to provide useful contextual information to the user in an unobtrusive way.

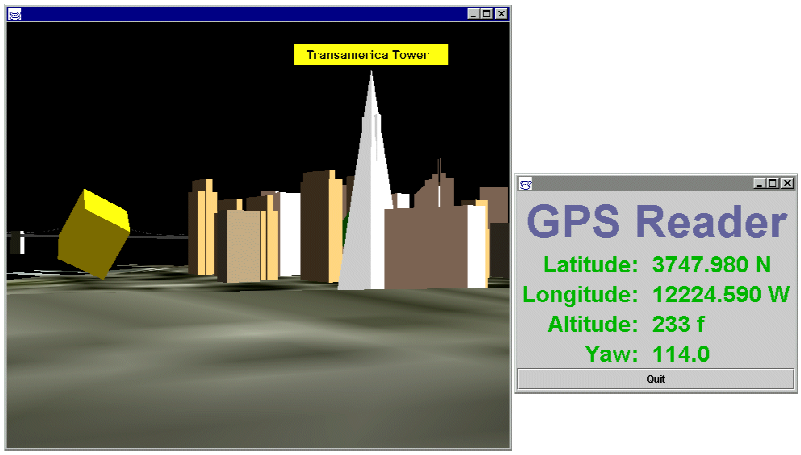

Figure 2: TravelMATE, easy and natural access to touristic information

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 emphasize that future interfaces will be more readily adopted by the population of users if they are simple, natural, intuitive, and familiar.

Within CANOES, we are also working on an "OfficeMATE" prototype to investigate how AR could enhance the workplace. Since one main focus is to take advantage of the proliferation of sensors and recognizers in enhanced spaces, we are continuing our efforts in data fusion, extending work produced in the MAESTRO project [7]. Other domains we are investigating include smart homes, smart appliances, and stock monitoring. For each prototype created, we will look at solutions for problems such as information sharing, privacy, communication quality, and so forth.

References

[1] A. Cheyer and L. Julia. Multimodal Maps: An Agent-based Approach. In book *Multimodal Human-Computer Communication, Lecture Notes in Artificial Intelligence #1374*, Springer, 1998.

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

[3] D. Guzzoni, A. Cheyer, L. Julia and K. Konolige. Many Robots Make Short Work. Report of the SRI International mobile robot team at AAAI96. *AI Magazine*, Spring 1997.

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

[5] L. Julia. Tasking Robots through Multimodal Interfaces: the "Coach Metaphor". In book *Collective Robotics, Lecture Notes in Artificial Intelligence #1456*, Springer, 1998.

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

[7] Z. Rivlin *et al*. MAESTRO: Conductor of Multimedia Analysis Technologies. *Communications of the ACM*, to appear in 1999.

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