Perceptive Assistive Agents in Team Spaces

Lisa D. Harper, Abigail S. Gertner, James A. Van Guilder
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102-7508
+1 703-883-5241
lisah@mitre.org

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces – *natural language, evaluation/methodology, interaction styles.*

General Terms
Design, Experimentation, Human Factors.

Keywords
perceptual computing, assistive agents, intelligent room.

1. INTRODUCTION
The use of next generation interface technologies to facilitate human-human collaboration and human-computer interaction (HCI) has tremendous potential for enhancing team interaction and performance in collaborative environments. However, our experience has shown us that the insertion of new devices and technologies cannot be accomplished successfully without 1) adapting to the dynamics of group interaction and 2) facilitating process change in the context of technological change. The addition of next generation capabilities and devices does not necessarily solve problems of communication and team collaboration. For example, though interactive software "whiteboards" are developed and marketed as groupware tools that facilitate group brainstorming, how much do we actually use them in day-to-day meeting contexts? Despite the availability of all of these new tools, computational systems are still perceptually impoverished in terms of understanding natural capabilities of humans including communication, motor and perceptual skills. They do not take into account the ways in which humans naturally interact with each other and with the world.

2. PERCEPTIVE AGENTS
Perceptual computing is concerned with using information from a variety of sensors such as video, audio and touch to make the computer aware of what a human is saying or doing via head, hand or body gestures. It also encompasses multimodal input and output. Though team spaces are supported by a variety of technologies from infrastructure to applications to devices, the combined use of these technologies is limited by physical context awareness. That is, the environment is not very adaptable to changes in group structure since elements of the environment (for example, noise level, location and number of participants, focus of activity, user technology savvy, and availability of resources) are not taken into account. We focus on the ability of assistive agents to monitor, access, and manipulate elements of the physical context, such as locating speakers and adjusting audio input to enhance speech recognition and remote listening quality or enriching the collaborative experience of remote participants via sound localization, panoramic video and shared views. Our hypothesis is that through the use of perceptive team space agents managing interactions between humans and the collaboration environment we will enhance team interaction and performance in both local and remote group interaction.

Perceptual interfaces seek to leverage information about how humans interact in the physical world. Sensors and devices should be transparent and passive. Pentland [3] proposes that perceptual intelligence is key to interfacing with future generations of machines. These machines include both large-sized interfaces (e.g., large plasma displays) as well as wearable and other ubiquitous devices. Familiar WIMP (windows, icon, menus, pointer) interfaces are not ideally designed for these sorts of computing environments [9].

3. THE EXPERIMENTAL TEAM ROOM
Though we believe that perceptual environments and intelligent assistive agents may change the way we think about human computer interaction, HCI researchers to date have provided relatively little evidence to support this intuition. If we are to assess the potential of advanced HCI technologies to impact the collaborative process, we require a means to experiment on actual users in a working environment. The MITRE Experimental Team Room (ETR) provides a platform for experimentation.

We have two objectives: 1) enhance the effectiveness of assistive agents by leveraging perceptual information accessible from the physical environment and 2) provide an experimental foundation for evaluating sophisticated HCI concepts, specifically, for assessing the potential benefits of perceptive assistive agents in the context of team room interactions. These interactions range from small informal working group meetings, discussion groups, instruction, and presentation. We are not only concerned with team activity when participants are co-present, but are also concerned with teams including remote participants – especially those connecting from the desktop or with mobile devices.

We have implemented an initial ETR facility (see Figure 1) built on the existing state of the art MITRE team room facilities, which is envisioned as a future experimental facility for internal operational use of advanced collaboration technologies and human computer interfaces. Rather than being a static showcase of cutting edge technologies, it is seen as the first step in a "plug-

IUI'04, Jan. 13–16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

and-play" user experimentation lab where technology and operational use emerge in an ongoing manner.

Figure 1: 3D model of the ETR

We are specifically assessing the use of this technology as it relates to team performance of actual users of a collaborative environment. We focus on metrics of productivity, user satisfaction, effectiveness as well as more objective component-level measures such as word error rate in speech recognition. The ETR provides not only a very natural and rich environment for this kind of research but also has the means for the instrumentation and observation of group activities. There is a glass observation deck and adjacent server room.

To date, we have constructed an "alpha" capability in the ETR. We installed two 50" plasma displays, one of which has a touch-sensitive overlay. We have tested the use of Bluetooth and other wireless devices such as user input devices (e.g., touch pads, pointers and keyboards). A prototype embodied conversational agent, Emma (Electronic MITRE Meeting Assistant) interfaces with the room's control system and provides the comprehensive set of functions currently available from the touchpad control panel sitting in every team room. We have also created a distributed software interface that enables Emma to control applications on team room PCs. Users can communicate with Emma either via speech or through a context-sensitive graphical user interface (see Figure 2).

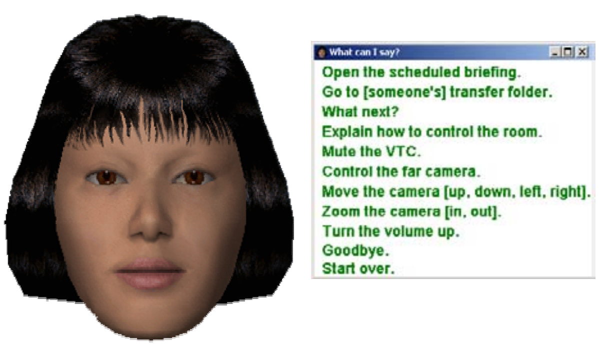

Figure 2: Emma's head and "What Can I Say" window

Currently, Emma is able to control lighting, speaker volume and camera angles, turn on the displays, initiate video-teleconference connections, and navigate PowerPoint presentations. She can also access the corporate network to retrieve files and look up basic employee data such as phone number and user ID. Emma has both tutorial and help modes to assist all levels of potential users. She was designed using custom head animation software, COTS automatic generation of synthetic speech and a research dialogue engine from Mitsubishi Electronic Research Laboratory [5]. Ultimately, Emma provides an alternative interface to the current touchpad interface and will gradually be equipped to deal with a greater number of resources as well as communicate with other "Emmas" residing on desktops interfacing with other team rooms.

4. INTERACTION IN THE ETR

It has frequently been observed (and is a source of great frustration to many) that setup in our current team rooms requires at least ten minutes at the start of every meeting to get the correct device inputs displayed on the screen, make necessary video teleconference connections, etc. Our first goal for Emma, therefore, was to replace the current touchpad room control interface with an assistive agent who could help users achieve these setup tasks quickly and straightforwardly.

Currently, interaction with Emma is task-oriented and primarily user-initiative, although Emma can suggest actions for herself or the user to perform based on task context. Figure 3 illustrates a typical dialogue with Emma.

User: Let's present a briefing

Emma: Would you like me to prepare the room for presenting a briefing?

User: Yes

Emma: *Turns down the lights, turns on the computer screen, sets the screen input to the room PC, starts PowerPoint running on the room PC.*

Emma: I finished preparing the room for presenting a briefing.

What would you like to do next?

Figure 3: Sample dialogue interaction with Emma

Though Emma has great promise as a meeting assistant who manages devices in the room and accesses scheduling and other information for users, she is effectively "blind" and cannot see user actions except for very specific sorts of events. To remedy this, we are in the process of creating a sensor-based environment, focusing particularly on microphone, video, wireless resource discovery, and identity detection (e.g., via RF emitters on badges). With these sensors in place, Emma will have access to information about *who* is in the room, *where* they are, *what* they are doing, and *how* they are doing it. She will use this information to enhance interaction in a number of ways.

As an embodied agent, information about Emma's focus of attention and state are visual. That is, users can see if she is "awake" and listening. Given access to perceptual information such as speaker location, she will be able to integrate her physical activity more closely with the activity in the room. For instance, she can turn her head to look at different speakers, gesture toward people in the room, and use non-verbal cues to ground the conversation.

Emma's knowledge of who is present can be used to automatically configure the room according to the preferences of those users, and to pre-fetch relevant resources (e.g. the meeting participants' shared network folders). Combining this with scheduling information, Emma can make the appropriate video teleconference connection and open the relevant PowerPoint file as soon as the meeting participants enter the room. If meeting participants are absent, Emma should be aware of this and be able to contact them upon request.

Another use of perceptual information is to dynamically improve audio quality. Current team rooms suffer from poor audio quality across remote connections. Speech recognition is also poor in these environments. A significant factor is the presence of multiple sources of noise in the environment including other speakers. Even if there were only one speaker, that speaker is often mobile. We believe that a network of phased array microphones that is able to identify the most important signal in the room and dynamically filter out other noises may provide the key to this puzzle. This ability to identify the most relevant source of auditory information and focus on it is not too difficult for human participants, but very challenging for machines.

Finally, we are concerned with enhancing a sense of "near presence" among remote meeting participants. Ultimately, teams have members that are not physically co-present at some time or another. Often they connect in via VTC or telephone. But they typically don't feel as included in the interaction as if they were actually present. Emma's perceptual information can be leveraged to support remote users sense of co-presence in a variety of ways. For instance, a panoramic camera that automatically adjusts its position to identify shared visual focus of attention might be desirable. But recognizing that individuals also maintain an individual focus of attention, remote users should be able to manipulate visual attention manually as well.

5. RELATED WORK

There are a variety of perceptive user interfaces whose research focus intersects with Emma. Location-aware devices have been used for tour guidance providing user context-sensitive information based on where a user is and has been previously [7],[8]. The Olivetti Active Badge [10] has been used to dynamically update the telephone extension nearest to a user. The Remembrance Agent [4] has been used to provide information relevant to the user's context, for example class notes when entering a specific classroom. Similarly, perceptive interfaces have been used for recognizing gaze, head and hand gestures for focus of attention, acknowledgement, or pointing [1],[6].

The most analogous research in group environments are Smart Room or Intelligent Room interfaces like those developed at the MIT AI Lab [2], Fuji Xerox Labs and NIST. The primary difference between our work and what other researchers have accomplished in this area is our integration of a perceptive environment with an embodied conversational agent. Emma serves as the "face" of the ETR and can use dialogue and gesture to guide users through the performance of complex tasks. By actively monitoring changes in the environment Emma can interact with meeting participants to ensure that the room continues to support team interaction as effectively as possible.

6. CONCLUSION

Humans are very efficient at processing a vast amount of information in the physical environment. This includes information that we can perceive via touch, hearing and vision. Traditional computing approaches make little use of this kind of information. We believe advanced HCI technologies should and will make more use of this information so that interfaces are both more contextually aware and adaptive to the way humans naturally communicate and interact. Not only is some of this technology now ripe enough to harvest by collaboration researchers and also those concerned with building assistive interfaces, but is on the horizon for commercial use.

We believe that pursuing work with Emma as a perceptive assistive agent, one that detects and reacts to changes in dynamic meeting environments, will ultimately find natural application in real-world collaborative settings. Critically, we don't desire to design yet another flashy HCI technology that *seems* useful, but to develop and assess the effectiveness of the use of an instantiation of particular operational concepts in an actual working environment. By leveraging our own operational mission we have a tremendous opportunity to study real user populations over an extended period of time in such a way as to inform usability issues critical to real collaborative environments.

7. REFERENCES

[1] Davis J., and Vaks, S. A Perceptual User Interface for Recognizing Head Gesture Acknowledgements, *ACM Workshop on Perceptual User Interfaces*, Orlando, Florida, November 15-16, 2001.

[2] Hanssens, N., Kulkarni, A., Tuchinda, R. and Horton, T. Building Agent-Based Intelligent Workspaces. In *ABA Conference Proceedings*. June 2002.

[3] Pentland, A. Perceptual Intelligence, *Communications of the ACM*, March 2000, 43-3.

[4] Rhodes, B. "The Wearable Remembrance Agent: a system for augmented memory." *Proceedings of the International Symposium on Wearable Computing,* IEEE, 1997.

[5] Rich, C., Sidner, C.L., Lesh, N.B., "COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction", *Artificial Intelligence Magazine*, Winter 2001 (Vol 22, Issue 4, pps 15-25)

[6] Ruddarraju R., Haro A., Essa I., Fast Multiple Camera Head Pose Tracking, In *Proceedings, Vision Interface 2003*, Halifax, Canada.

[7] Smailagic, A. and Martin, R. "Metronaut: A Wearable Computer with Sensing and Global Communication Capabilities". *Proceedings of the International Symposium on Wearable Computing,* IEEE, 1997.

[8] Sumi, Y., Etani, T., Fels, S., Simone, N., Kobayashi K. and Mase, K. "C-MAP: Building a Context-Aware Mobile Assistant for Exhibition Tours". *Social Interaction and Community ware*, Japan, June 1998.

[9] Turk, M. and Robertson, G. Perceptual user interfaces (introduction), *Communications of the ACM*, March 2000, 43-3.

[10] Want, R., A. Hopper, V. Falcao, and Gibbons, J. "The active badge location system". *ACM Transactions on Information Systems*, 10(1): pp. 91-102, January 1992.