

# Interacting with Groups of Computers

*AUIs recognize human attention in order to respect and react to how users distribute their attention in technology-laden environments.*

AS WE EVOLVE NEW RELATIONSHIPS WITH THE COMPUTING systems that surround us, there is a continuous need to adopt new strategies for user interface design. Many of the features of the graphical user interface (GUI) were designed under the assumption that computers would be used as isolated tools with a one-to-one relationship with

By Jeffrey S. Shell, Ted Selker, and Roel Vertegaal

users. But today, each user has many computers, causing existing channels of interaction to break down. The reason for this is that computers have no knowledge of the devices or tasks a user is attending to. As a consequence, users are bombarded with interruptions from their PDAs, email programs, instant messaging applications, and cell phones. The nature of these interruptions is often acute, demanding full and immediate attention.

To design less intrusive and more sociable interfaces, we suggest augmenting computing devices with attention sensors that allow the devices to prioritize their demands for user attention. Thus, users and devices may enter a turn-taking process similar to what naturally occurs in a human group conversation. This process is key to a new paradigm for computer interfaces—Attentive User Interfaces (AUIs). Here, we present some of the prototype AUIs designed at Queen's University and MIT. We describe scenarios demonstrating how to design systems that engage users in a manner complementary and appropriate to their attentive context, in order to improve interactions among people and ubiquitous computers.

People communicate attention to each other all the time. Gestures, looks, laughs, and other nonverbal utterances often serve to stimulate the

listener, making conversations more interesting and engaging. However, nonverbal cues communicate more than just attention. While eye contact is a powerful communicator of attention between people, too much of it can make us uncomfortable and too little leaves us feeling ignored. Like this example shows, nonverbal communication of attention is always interpreted in context. By viewing attention in a social context, we can design systems able to engage in richer, more meaningful interactions with people. AUIs allow user attention to drive the human-computer interface scenario in physical and virtual environments. By recognizing attentive cues from users, and by communicating attention to users, these interfaces encourage a more natural process of turn taking.

All interfaces use some method to negotiate control between computer and user. When



computers do not follow reasonable conventions for flow of control, they generate interruptions that are intrusive and annoying. Consider the example of the email tool in Figure 1, which brings up a modal dialogue box to inform the user that a message has been received. Without any regard for the user's current activity, the dialogue box pops up in the center of the screen. The user can continue his or her activities only by clicking the "OK" button. This example points to a serious underlying flaw in current user interfaces—their lack of knowledge of a user's current activities. This problem is intensified because users are now surrounded by many computer systems, each competing for the user's attention. This scenario is analogous to human group communication, in which many people might simultaneously have an interest in speaking.

Clearly, human attention is a limited resource in conversations. A person can only listen to, and fully absorb, the message of one individual at a time. When there are many speakers, the Cocktail Party Effect allows us to focus on the one person we are interested in by attenuating speech from other individuals. However, a more effective method to regulate group communication is to have speakers take turns. According to Short et al. [10], as many as eight cues can be used to negotiate conversational turn taking. Of these, only eye gaze allows people to continuously perceive who is paying attention to whom. We found that visual attention conveyed by eye contact is a reliable indicator of whom one speaks to or listens to during group conversations. It is also a social cue that conveys when it is time for a speaker to relinquish the floor, and who is expected to speak next [1]. Eye contact functions as a nonverbal visual signal that peripherally conveys attention without interrupting the verbal auditory channel. With it, humans achieve a remarkably efficient process of conversational turn taking. Without it, turn tak-

ing breaks down [11].

To facilitate turn taking between devices and users in a nonintrusive manner, AUIs monitor nonverbal attentional channels, such as eye gaze,



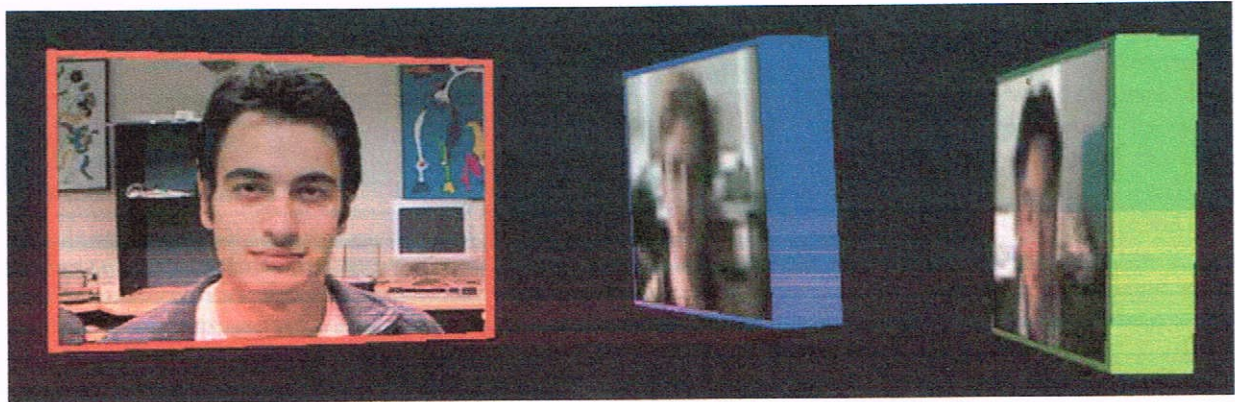
Figure 1. Email application with modal notification alert.

to determine when, whether, and how to communicate with a user. Devices that negotiate requests for attention over peripheral channels make human-device communication more efficient, reliable, and sociable.

## Goals of AUIs

AUIs aim to recognize a user's attention space in order to optimize the information-processing resources of user and devices. This is accomplished by measuring and modeling the users' past, present, and future attention for tasks, devices, or people. Key features of AUIs include:

- *Sensing attention.* By monitoring users' physical proximity, body orientation, and eye fixations, AUIs can determine what device, person, or task the user is attending to.
- *Reasoning about attention.* By modeling user attention, AUIs can estimate task prioritization and predict attentive focus.
- *Graceful negotiation of turns and sense user acknowledgment of the request.* Before taking the foreground, AUIs determine whether the user is available for interruption given the priority of the request; signal the user via a nonintrusive peripheral channel; sense user acknowledgment of the request.
- *Communicating attention.* To encourage efficient turn taking, AUIs communicate their attention to users, and communicate the



attentive focus of the user to other AUIs and remote people that request the user's attention.

- *Augmenting attentive resources.* Analogous to the Cocktail Party Effect, AUIs may optimize the use of the user's attentive resources by magnifying information in the estimated focus of user activity, while attenuating peripheral detail.

### Previous Work

Rick Bolt's Gaze-Orchestrated Dynamic Windows [2] was one of the first true AUIs. It simulated a composite of 40 simultaneously playing television episodes on one large display. All stereo soundtracks from the episodes were active, creating "a kind of Cocktail Party Effect mélange of voices and sounds." Via a pair of eye-tracking glasses, the system sensed when the user looked at a particular image, turning off the soundtracks of all other episodes. If users looked at one episode for a few seconds, the system would zoom in to fill the screen with that image. Because eye movements are not always voluntary, they are best interpreted as an indicator of interest, rather than as a means for control. Similarly, Nielsen's Noncommand Interfaces [8] observed user activity and reacted to implicit input based on simple, predefined heuristics, instead of responding to explicit, user-issued commands (for example, mouse clicks).

Vertegaal's GAZE [12] was one of the first AUIs to apply the Noncommand principle to communicate user attention during remote, collaborative interactions. Using eye trackers, GAZE observes whom and what participants look at during mediated group conversations (Figure 2). By automatically rotating 2D video images of individuals toward the person they look at, participants in a 3D meeting room can see

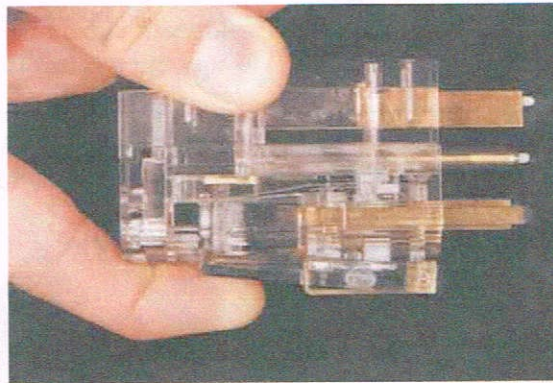
who is talking to whom. **Figure 2. GAZE-2 attentive videoconferencing.**

According to Maglio et al., not only do users look at other people when speaking to them, they also look at the devices that execute spoken commands [6]. This means a person's eye gaze can be used to open and close communication channels with devices. We

applied this principle in the design of several AUIs described later. However, it is important to note that user attention can be observed through many means besides eye tracking. With Priorities [3], Horvitz designed the first AUI to forward a user's email messages to digital appliances on the basis of their perceived urgency. Messages are prioritized using simple measures of user attention to a sender:

the mean time and frequency with which the user responded to email messages from that sender. Messages with a high priority rating are forwarded to a user's pager, while messages with low priority can be checked at the user's convenience.

Similar in nature to AUIs, Context-Aware Systems [5, 9] employ the user's physical situation, goals, and experience, as well as the system's capabilities to inform action. These systems can recognize and handle repetitive, work-intensive subtasks to allow users to do less to accomplish their goals. Unlike AUIs, user attention is not the primary criterion to determine user context. For example, the Universal Plug (Figure 3) is a tool capable of functioning in several contexts, without any knowledge of the user's activities. When the plug is pressed against a power outlet anywhere in the world, it automatically selects the correct power and voltage. The correct prongs enter



**Figure 3. Context aware, not attentive.**

the outlet, while the others retract without any user intervention. Being a tool, the plug does not vie for user attention, thus the attentive status of the user is not required to use the plug. The difference between AUIs and Context-Aware Interfaces is that context is always dominated by user attention in an AUI framework.

### Prototypes that Sense Attention

Here, we introduce some of the prototypes recently developed at Queen's University and MIT. We begin our discussion by presenting novel attention sensors. To enable a seamless turn-taking process between humans and groups of

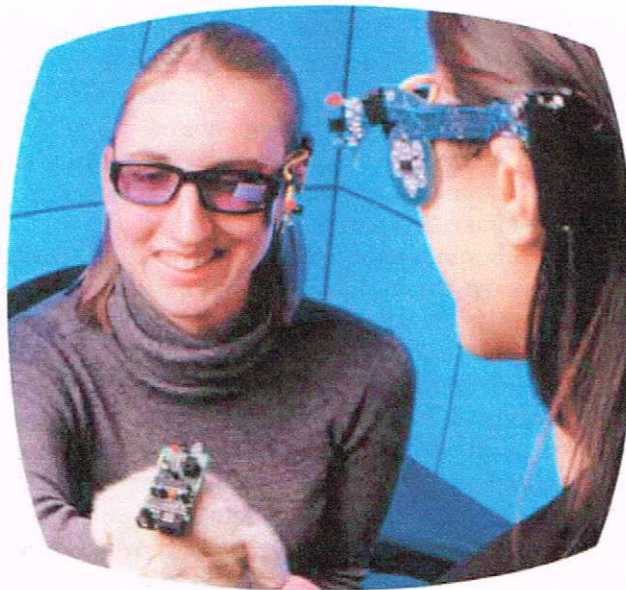


Figure 4. Eye aRe glasses

the sensor. Unlike most commercially available eye trackers, eyeCONTACT is inexpensive, unobtrusive, tolerant to user head movement, and requires no calibration.

aRe devices. Software determines when the user blinks in order to detect aspects of the user's cognitive load, for example, stress and fatigue levels.

Our second attention sensor, eyeCONTACT (Figure 5a), is based on the IBM PupilCam [7]. It consists of a camera that uses computer vision to find pupils in its field of view and detect when users look at

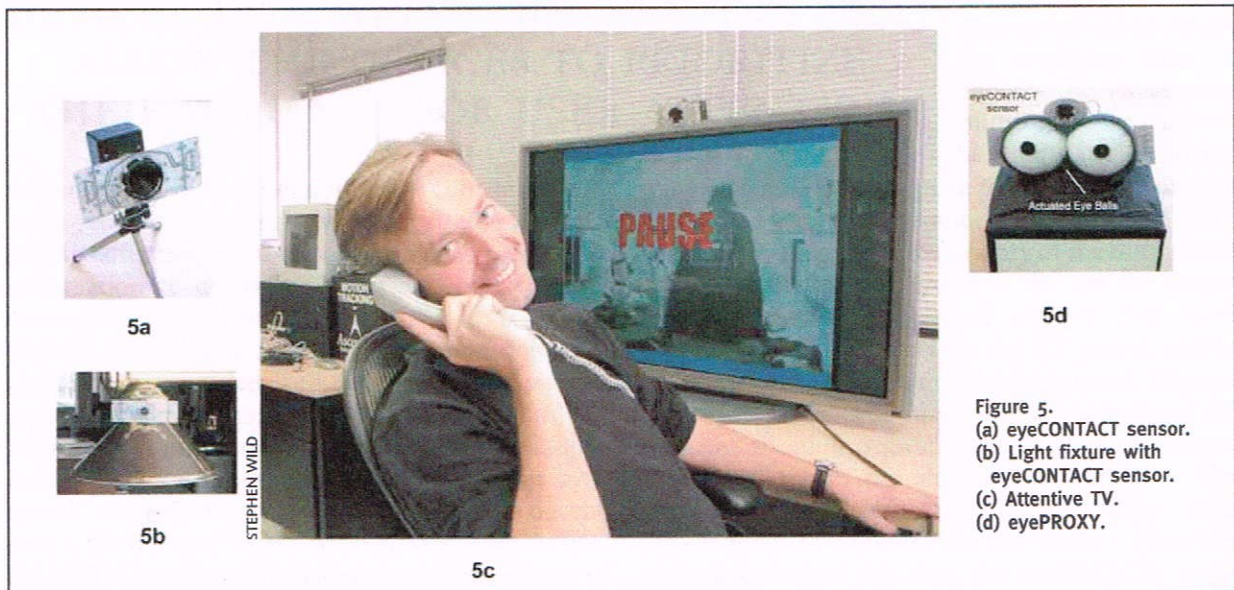


Figure 5. (a) eyeCONTACT sensor. (b) Light fixture with eyeCONTACT sensor. (c) Attentive TV. (d) eyePROXY.

computers, devices must also communicate attention for the user. Using scenarios, we will illustrate the application of attention sensors in appliances that reason about attentive input and, in turn, convey their own attention.

The first attention sensor is Eye aRe (Figure 4), a simple eye movement detection system. Eye aRe glasses report whether the user is looking in the direction of another device or user, augmented with Eye aRe capabilities. Eye aRe detects both pauses in the user's eye movements and light emitted from other Eye

By embedding eyeCONTACT sensors in household appliances and other digital devices we designed eyePLIANCES, which explore gradual turn taking between humans and attentive appliances. By looking at an eyePLIANCE a user conveys attention for the device, which is used to regulate communications. A user interacts with the device with speech commands, or by using remote or manual controls. Figure 5b shows the simplest form of an eyePLIANCE, an attentive light fixture. A user can switch the light on or off by simply saying "on" or "off" while looking at the fix-

ture. By having only one device listen at a time, speech recognition is simplified as generic terms such as “on” and “off” can be reused for different devices. Our experiences indicate that eyeCONTACT sensors, as pointing devices for the real world, make it easier to communicate the target of remote interactions.

### Negotiating User Attention

In environments with many attention-sensing appliances, AUIs need a dynamic model of the user’s attentive context to establish a turn-taking process. This context includes which task, device, or person the user is paying attention to, the importance of that task, and the preferred communication channel to contact the user. eyeREASON is a personalized communications server that negotiates all remote

interactions between a user and attentive devices by

## TO DESIGN LESS INTRUSIVE AND MORE SOCIABLE INTERFACES, WE SUGGEST AUGMENTING COMPUTING DEVICES WITH ATTENTION SENSORS THAT ALLOW THE DEVICES TO PRIORITIZE THEIR DEMANDS FOR USER ATTENTION.

keeping track of the user’s attentive context. Appliances report to the server when they sense a user is paying attention to them. eyeREASON uses this information to determine when and how to relay messages from appliances to the user. This is accomplished using knowledge of what communication channels are occupied, and the priority of the message relative to the tasks the user is engaged in [3]. All speech communication between user and appliances is processed through a wireless headset by a speech recognition and production system on the server. As the user works with various devices, eyeREASON switches its vocabulary to the lexicon of the focus device, sending commands through that device’s I/O channels.

The following scenario illustrates interactions of a user with various eyePLIANCES through eyeREASON. It shows how an awareness of the user’s attentive context may facilitate graceful turn taking between users and remote ubiquitous devices.

Alex enters his living room, which reports his presence to his eyeREASON server. He turns on his television, which has live-pausing capability (Figure 5c). The television is augmented with an eyeCONTACT sensor, which notifies the server that it is being

watched. The eyeREASON server updates the visual and auditory interruption levels of all people present in the living room. Alex goes to the kitchen to get himself a cold drink from his attentive fridge, which is augmented with a radio tag reader. As he enters the kitchen, his interruption levels are adjusted appropriate to his interactions with devices in the kitchen. In the living room, the TV pauses because its eyeCONTACT sensor reports that no one is watching. Alex queries his attentive fridge and finds there are no cold drinks within. He gets a bottle of soda from a cupboard in the kitchen and puts it in the freezer compartment of the fridge. Informed by the radio tag on the bottle, the fridge estimates the amount of time it will take for the bottle to freeze and break. It records Alex’s tag and posts a notification with a timed priority level to his eyeREASON server.

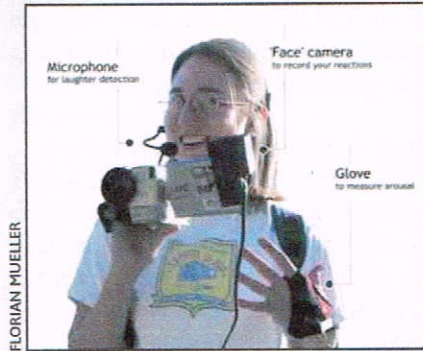
Alex returns to the living room and looks at the TV, which promptly resumes the program. When the notification times out, Alex’s eyeREASON server determines the TV is an appropriate device to use for notifying Alex. It chooses the visual communication channel, because it is being

watched and is less disruptive than audio. A box with a message from the fridge appears in the corner of the TV. As time progresses, the priority of the notification increases, and the box grows in size on the screen, demonstrating with increased urgency that Alex’s drink is freezing. Alex gets up, the TV pauses, and he sits down at his computer to check his email. His eyeREASON server determines that the priority of the fridge notification is greater than that of his current email, and moves the alert to his computer. Alex acknowledges this alert, and retrieves his drink, causing the fridge to withdraw the notification. Had Alex not acknowledged this alert, the eyeREASON server would have forwarded the message to Alex’s email, instead of continually notifying him directly.

### Communicating Device Attention

To enable efficient and sociable interactions between users and devices, attentive systems must, conversely, convey their attention to a user. Figure 5d shows how eyePLIANCES may communicate their own attention using an eyePROXY. An eyePROXY consists of an eyeCONTACT sensor mounted on a pair of actuated, moveable eyeballs. It can be connected to any eyePLIANCE to provide nonverbal feedback

## Students "Sense" AUI Solutions

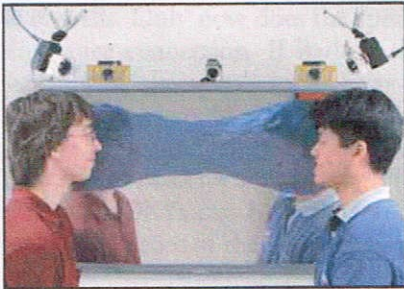


### LAFCam

The LAFCam makes use of the involuntary attentive cues people utter. We trained an AI model to recognize MIT Media Lab student Andrea Lockerd's laugh and voice. We recorded Andrea walking around Harvard Square making a videotape using the system. LAFCam was then able to find and mark the three most engaging moments in the video based on the nonverbal utterances Andrea inadvertently made while filming, highlighting points of interest from her perspective to the audience.

### AuraMirror

AuraMirror is a media art project by Human Media Lab student Alexander Skaburskis. The video mirror renders the virtual windows of attention, or attentive auras, that encompass groups of people during interactions. A visual representation of conversational attention is obtained by superimposing bubbles over each participant's head. These 'auras' grow toward interlocutors to form tunnels during sustained interactions. This permits users to see how they distribute their attention in group interactions, and the effect of interruption on this process. When interlocutors look at the mirror to see their merged aura, it will invariably break, because the target of their visual attention has changed. This serves as a metaphor for interruption.



### Social Floor



The floor in the Context Aware Computing Lab at MIT senses people's position and uses this information to comment on social relationships. The reasoning is based on the common-sense notion that proximity is related to interest. People are probably attending to other people and objects that are close by. This simple AUI notices social distance and paints butterflies around the feet of people standing next to each other. If one group of people is standing separate from another, the floor projects footsteps between them. This is done to encourage participants to notice the social distance and possibly attend to each other. If one person is standing apart from a group in a designated location, it projects a podium and activates a spotlight on the speaker. If a lone visitor stands near a demo, a cartoon head appears on the floor, describing the project using localized speakers. This scenario shows that even a crude metric of position can be used to deduce aspects of what a person is attending to.



### Attentive Cell Phone

To avoid the problem of phones interrupting face-to-face conversations in public places, Human Media Lab student Connor Dickie (left) augmented a cell phone with a wearable eyeCONTACT sensor. The eyeCONTACT sensor reports when someone looks at Connor. The attentive cell phone uses this information to assess whether Connor is engaged in a conversation. The cell phone communicates his attentive status to people in his customized contact list. If Connor is not available, a picture of the back of his head is displayed. If a message is urgent, callers can override Connor's preferred method of notification, which is currently set to

vibrate. By adding a nonintrusive channel to convey attention, the attentive cell phone encourages social behavior among users, thus reducing the number of interruptive phone calls one receives. **C**

to the user, demonstrating this appliance is now listening, or requesting a turn. An eyePROXY may also serve as a surrogate that indicates the attention of a remote individual [4]. We augmented a speakerphone with an eyePROXY to experiment with gradual negotiation of communications using non-verbal channels. The following scenario illustrates the process.

Arnie wishes to place a call to Barbara. He looks at Barbara's speakerphone proxy on his desk, which detects eye contact and begins setting up a voice connection with Barbara. On the other side of the line, Arnie's proxy on Barbara's desk starts moving its motorized eyeballs, using its eyeCONTACT sensor to find Barbara's pupils. Barbara observes the activity of Arnie's proxy in her peripheral vision, and looks at the eyeballs. Only now does the speakerphone establish a voice connection. If Barbara does not wish to take the call, she simply looks away from the proxy. Barbara's proxy would then convey her unavailability to Arnie by shaking its eyes, and breaking eye contact. To avoid the need for multiple eyePROXYs per location, eyePROXYs can be augmented with a display showing a picture of the current caller.

## Discussion and Outlook

The popularity of ubiquitous, wireless computing devices has fundamentally changed the way we interact with technology. We feel it is necessary to augment devices with attention-sensing capabilities to help users manage the many conflicting requests for their attention. Sensing technology has improved in cost and functionality to the extent we can now reliably monitor users to determine what they are paying attention to. AUIs may measure attention in many ways. In social settings, the physical distance between people, the way they turn their heads, and the way they direct their eye gaze at each other all indicate attention. Obtaining nonverbal attentional cues and using them in context allows us to build systems that respectfully and efficiently manage a user's attention space. This permits more natural, sociable, and most importantly, meaningful interaction between people and groups of computers. We have presented a series of systems and scenarios that describe how we approach this problem. As designers, however, we must keep in mind socio-technological issues that may arise from the usage of attentive systems. For instance, will people trust a technological system to serve as the gatekeeper to their interactions? How can we foster such a trust, and safeguard the privacy of the people using systems that sense, store, and relay information about the

user's identity, location, activities, and communications with other people?

## Conclusion

We have presented here an overview of our work on AUI—interfaces that recognize, refine, and respect a user's attention space. By augmenting devices and appliances with attention sensors that permit the devices to recognize and prioritize demands on the user's attention, users and devices may enter a turn-taking process analogous to that found in human group conversation. By explicitly designing for the virtual windows of attention between devices and users, interactions with groups of computers may become more sociable as well as more efficient. **□**

## REFERENCES

1. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. 1976. Cambridge University Press, London.
2. Bolt, R.A. Conversing with computers. *Technology Review* 88, 2 (1985), 34–43.
3. Horvitz, E., Jacobs, A., and Hovel, D. Attention-sensitive alerting. In *Proceedings of UAI '99 Conference on Uncertainty and Artificial Intelligence*. 305–313.
4. Greenberg, S., and Kuzuoka, H. Using digital but physical surrogates to mediate awareness, communication and privacy in media spaces. *Personal Technologies* 4, 1.
5. Lieberman, H., and Selker, T. Out of context: Computer systems that adapt to, and learn from, context. *IBM Systems J.* 39, 3&4 (2000), 617–632.
6. Maglio, P., Matlock, T., Campbell, C., Zhai, S., Smith, B.A. Gaze and speech in attentive user interfaces. In *Proceedings of the Third International Conference on Multimodal Interfaces*. (Beijing, China, 2000), 1–7.
7. Morimoto, C., Koons, D., Amir, A., and Flickner, M. Pupil detection and tracking using multiple light sources. *Image and Vision Computing* 18 (2000), 331–335.
8. Nielsen, J. Noncommand user interfaces. *Commun ACM* 36, 4 (Apr. 1993), 83–99.
9. Selker, T., and Burleson, W. Context-aware design and interaction in computer systems. *IBM Systems J.* 39, 3&4 (2000), 880–891.
10. Short, J., Williams, E., and Christie, B. *The Social Psychology of Telecommunications*. 1976. Wiley, London.
11. Vertegaal, R. and Ding, Y. Explaining effects of eye gaze on mediated group conversations: Amount or synchronization? In *Proceedings of CSCW 2002 Conference on Computer Supported Cooperative Work*. (New Orleans, Nov. 2002) ACM Press, NY, 41–48.
12. Vertegaal, R. The GAZE Groupware System: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of CHI '99 Conference on Human Factors in Computing Systems*. (Pittsburgh, 1999). ACM Press, NY, 294–301.

FOR A MORE EXTENSIVE LIST OF REFERENCES, SEE [WWW.HML.QUEENSU.CA/CACM-REFS.HTML](http://WWW.HML.QUEENSU.CA/CACM-REFS.HTML). THE AUTHORS ARE GRATEFUL TO THE MANY STUDENTS WHO HAVE CONTRIBUTED TO THE CONTEXT AWARE COMPUTING LABORATORY AND THE HUMAN MEDIA LABORATORY.

**JEFFREY S. SHELL** ([shell@cs.queensu.ca](mailto:shell@cs.queensu.ca)) is a graduate student with the Human Media Lab at Queen's University, Canada.  
**TED SELKER** ([selker@media.mit.edu](mailto:selker@media.mit.edu)) is a professor and director of the Context Aware Computing Laboratory at the MIT Media Lab.  
**ROEL VERTEGAAL** ([roel@acm.org](mailto:roel@acm.org)) is a professor and director of the Human Media Lab at Queen's University, Canada.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

© 2003 ACM 0002-0782/03/0300 \$5.00