Available online at www.sciencedirect.com

SCIENCE DIRECT®

Computers in Human Behavior 22 (2006) 771–789

Computers in Human Behavior

www.elsevier.com/locate/comphumbeh

Designing for augmented attention: Towards a framework for attentive user interfaces

Roel Vertegaal *, Jeffrey S. Shell, Daniel Chen, Aadil Mamuji

Human Media Laboratory, Queen's University, Kingston, Ont., Canada K7L3 N6

Abstract

Attentive user interfaces are user interfaces that aim to support the user's attentional capacities. By sensing the users' attention for objects and people in their everyday environment, and by treating user attention as a limited resource, these interfaces avoid today's ubiquitous patterns of interruption. Focusing upon attention as a central interaction channel allows development of more sociable methods of communication and repair with ubiquitous devices. Our methods are analogous to human turn taking in group communication. Turn taking improves the user's ability to conduct foreground processing of conversations. Attentive user interfaces bridge the gap between the foreground and periphery of user activity in a similar fashion, allowing users to move smoothly in between.

We present a framework for augmenting user attention through attentive user interfaces. We propose five key properties of attentive systems: (i) to sense attention; (ii) to reason about attention; (iii) to regulate interactions; (iv) to communicate attention and (v) to augment attention.

Keywords: Ubiquitous computing; Attentive user interfaces; Eye tracking; Notification; Context-aware computing

1. Introduction

It is our belief that the proliferation of ubiquitous digital devices necessitates a new way of thinking about human–computer interaction. Weiser (1991) said of ubiquitous computing: "The most profound technologies are those that disappear. They weave themselves

---

* Corresponding author.
E-mail addresses: roel@cs.queensu.ca (R. Vertegaal), shell@cs.queensu.ca (J.S. Shell), chend@cs.queensu.ca (D. Chen), mamuji@cs.queensu.ca (A. Mamuji).

0747-5632/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2005.12.012

into the fabric of everyday life until they are indistinguishable from it''. Although we live in a ubiquitous computing age, with technologies that have interwoven themselves with our daily existence, the interface has far from disappeared. For many years, our design and research efforts have focused on the development of computers as tools, extensions of analog devices such as paper, pencils and typewriters. While this view will continue to be extremely successful for years to come, we are now beginning to see limits to this approach. One of the main reasons for this is that, unlike traditional tools, computers are becoming increasingly *active* communicators. However, they are ill equipped to negotiate their communications with humans. Consider the example in Fig. 1. An email tool brings up a modal dialog box to inform its user that a message has been received. Without *any* regard for the user's current activity, the dialog box pops up in the *center* of her screen.

Only by clicking the ''OK'' button can the user continue her activities. This example points out a serious underlying flaw in user interfaces: the computer's lack of knowledge about the present activities of its user. Indeed, the behavior of such devices could be described as socially inadequate.

2. HCI as multiparty dialogue

As we evolve new relationships with the computing systems that surround us, there is a need to develop new strategies for design. We have moved from many users sharing a single computer through a command line interface, to a single person using one computer with a graphical user interface (GUI). Recently, we have developed a *multiparty* relationship with our computers, one that causes existing channels of interaction to break down because:

- Each user is surrounded by *many* active computing devices.
- These devices form part of a worldwide, connected network.
- Users form part of a worldwide ''attention seeking'' community through these active devices.

Because of the ubiquity of active connected devices, users are now bombarded with interruptions from their Palm Pilots, BlackBerries, email programs, auction trackers, instant messaging tools and cell phones. Like the pop up email example in Fig. 1, the nature of interruptions is often acute, requiring immediate attention. As a consequence, user attention has become a limited resource, continually vied for by various devices each

You have new mail.

Fig. 1. E-mail application with modal ''you have new mail'' notification alert.

claiming a high priority. We must design computers with channels to explicitly negotiate the volume and timing of communications with their user, pending the user's current needs. Our design strategy, to solve this problem by making interfaces more considerate (Gibbs, 2005) and less interruptive, rests upon the most striking parallel available: that of multiparty dialogue in *human group communication*.

3. Taking turns for attention

We were in part motivated by work performed in the area of social psychology towards understanding the regulation of human multiparty communication and attention. In human conversation, attention is inherently a limited resource. Humans can only listen to, and absorb the message of one person at a time (Cherry, 1953; Duncan, 1972). Thus, we have developed two attentive mechanisms to focus on a single verbal message stream: (1) When there are many speakers, the *Cocktail Party Phenomenon* allows us to focus on the words of the one speaker we are interested in by attenuating speech from other individuals (Cherry, 1953). We can apply this approach to augment the attentive capacities of users. (2) However, a far more effective method to optimize attention is to allow only one person to speak at a time, while the others remain silent. By using nonverbal cues to convey attention, humans achieve a remarkably efficient process of speaker exchange, or *turn taking* (Duncan, 1972). Turn taking provides a powerful metaphor for the regulation of communication with ubiquitous devices. So what information do humans use to determine when to speak, or yield the floor? According to Short, Williams, and Christie (1976), as many as eight cues may be used: completion of a grammatical clause; a socio-centric expression such as 'you know'; a drawl on the final syllable; a shift in pitch at the end of the clause; a drop in loudness; termination of gestures; relaxation of body position and the resumption of eye contact with a listener. However, in group conversations only one of these cues indicates to *whom* the speaker may be yielding the floor: eye contact (Vertegaal, 1999).

4. Eyecontact points to targets of attention

Eye contact indicates with about 82% accuracy whether a person is being spoken or listened to in four-person conversations (Vertegaal, Slagter, Van der Veer, & Nijholt, 2001). When a speaker falls silent, and looks at a listener, this is perceived as an invitation to take the floor. Vertegaal (1999) showed that in triadic mediated conversations, the number of turns drops by 25% if eye contact is not conveyed. According to a recent study, 49% of the reason why someone speaks may be explained by the amount of eye contact with an interlocutor (Vertegaal & Ding, 2002). Humans use eye contact in the turn taking process for four reasons:

1. Eye fixations indicate most reliably the target of a person's attention, including their conversational attention (Argyle & Cook, 1976; Vertegaal et al., 2001).
2. The perception of eye contact increases arousal, which aids in proper allocation of brain resources, and in regulating inter-personal relationships (Argyle & Cook, 1976).
3. Eye contact is a *nonverbal visual* signal, one that can be used to negotiate turns without interrupting the *verbal auditory* channel.
4. Eye contact allows them to observe the nonverbal responses, including the attentional focus, of others.

We have sought to implement similar characteristics in computing systems, in order to allow them to communicate more sociably with their users. The eye gaze of the user, as an extra channel of input, seems an ideal candidate for ubiquitous devices to sense their users' attention. It may allow devices to determine whether a user is attending to them, or to another device or person. By tracking whether a user ignores or accepts requests for attention, interruptions can be made more subtly.

5. Designing windows and mice for the real world

Bellotti et al. (2002) posed five challenges for multiparty HCI. In this paper, we hope to provide some suggestions towards answering the first three: (1) How do I address one of many possible devices; (2) How do I know the system is ready and attending to my actions; (3) How do I specify a target for my actions? How *do* we move from GUI–style interactions where multiple entities are represented on a single computing device to interactions with many remote devices in the real world? For one, it is important to note that many of the elements of GUIs were designed with attention in mind. According to Smith et al. (1982), *windows* provide a way to optimally allocate screen real estate to accommodate user task priorities. Windows represent foreground tasks at high resolution, and occupy the bulk of display space in the center of vision. *Icons* represent peripheral tasks at low resolution in the periphery of the user's vision. *Pointers* allow users to communicate their focus of attention to graphic objects. By clicking icons to open windows, and by positioning, resizing and closing windows, users use their pointing device to manually manage their attention space. By control-clicking graphic objects, users indicate the target of menu commands. In clicking "OK" buttons, users acknowledge interruptions by alert boxes. Fig. 2 shows how we might extend the above GUI elements to interactions with ubiquitous remote devices, drawing parallels with the role of attention in human turn taking. Windows and icons are supplanted by graceful increases and decreases of information resolution between devices in the foreground and background of user attention; Devices sense whether they are in the focus of user attention by observing presence and eye contact; Menus and alerts are replaced by a negotiated turn taking process between users and devices. Such characteristics and behaviors define an attentive user interface.

6. A framework for attentive user interfaces

Attentive user interfaces are interfaces that optimize their communication with users, such that information processing resources of users and devices are dynamically allocated

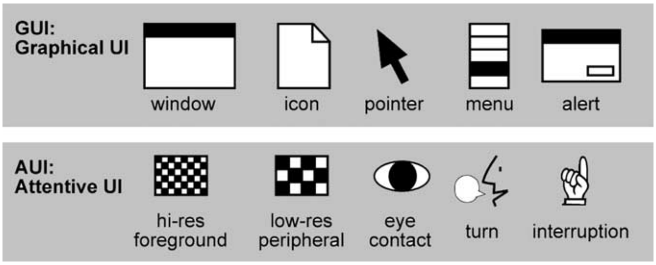


Fig. 2. Equivalents of GUI elements in Attentive UI.

according to the users' task priorities. This is achieved using measures and models of the users' past, present and future attention for tasks. Five key properties of AUIs include (Shell, Selker, & Vertegaal, 2003):

1. *Sensing attention*: By tracking users' physical proximity, body orientation and eye fixations, interfaces can determine what device, person, or task a user is most likely attending to.
2. *Reasoning about attention*: By statistically modeling simple interactive behavior of users, interfaces can estimate user task prioritization.
3. *Communication of attention*: Interfaces should make available information about the users' attention to other people and devices. Communication systems should convey whom or what users are paying attention to, and whether a user is available for communication.
4. *Gradual negotiation of turns*: Like turn taking, interfaces should determine the availability of the user for interruption by: (a) checking the priority of their request, (b) progressively signaling this request via a peripheral channel and (c) sensing user acknowledgment of the request before taking the foreground.
5. *Augmentation of focus*: The ultimate goal of all AUIs is to augment the attention of their users. Analogous to the cocktail party phenomenon, AUIs may, for example, magnify information focused upon by the user, and attenuate peripheral detail.

Modern traffic light systems provide an interesting parallel to an attentive user interface that augments the attention of all the users involved. They use presence sensors in the road surface to determine every vehicle's intent at the intersection, in effect *sensing the user's attention*. They are programmed with models that determine the priority of traffic on intersection roads with volume statistics, in effect allowing for *reasoning about attention*. Using peripheral displays, such as traffic lights, they *communicate the collective attention of drivers*. As such, they *negotiate turn taking* on intersections to allow for smooth traffic flow.

7. Other related work in AUIs

Our work was also inspired by our interactions with a host of researchers, designers and media artists, as well as by the vision of ubiquitous computing (Weiser, 1991) and the seamless interaction created by considering foreground vs. background in tangible user interfaces (Ishii & Ullmer, 1997). Since both of these paradigms are well known, we will limit ourselves here to a discussion of existing attentive user interfaces. We will discuss examples as they relate to our framework.

7.1. Sensing attention: eye tracking as a tool in AUIs

Rick Bolt's Gaze-Orchestrated Dynamic Windows was one of the first true AUIs (Bolt, 1985). 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 the user's visual attention towards a particular image, turning off the soundtracks of all other episodes and zooming in to fill the screen with the image. Bolt's system demonstrated how a windowing system could be translated into a display with malleable resolution that exploits the dynamics of the user's visual attention. It shows the great potential of AUIs to augment attention by reducing information overload in congested audio–visual environments. Jacob (1995) and MAGIC (Zhai, Morimoto, & Ihde, 1999) showed that eye tracking works best when it is applied to observe user attention, rather than as a device for control. This is because to the user, the eyes are principally an input rather than an output organ. As a consequence, when the duration of an eye fixation on an on-screen object is used to issue commands, users may unintentionally trigger unwanted responses while looking (The *Midas touch* effect (Jacob, 1995)). In a *non-command interface* (Nielsen, 1993) version, instead of a user explicitly issuing commands, the computer observes user activity. The system then *reasons about action* using a set of heuristics. In the classic game of Paddleball, the goal is to position a sliding paddle into the path of a moving ball using a joystick, which introduces an eye/hand coordination problem. In a non-command interface version of the game the paddle location is given by the horizontal coordinate of a user's on-screen gaze, communicating the visual attention of the user, and thus eliminating the game's eye/hand coordination problem.

7.2. Reasoning about attention

Attentional interfaces (Horvitz, 1999; Horvitz, Jacobs, & Hovel, 1999) are interfaces that use Bayesian reasoning to identify what channels to use and whether or not to notify a user. In the *Priorities* system (Horvitz et al., 1999), the delivery of email messages is prioritized using simple measures of user attention to a sender: the mean time and frequency with which the user responds to emails from that sender. Messages with a high priority rating are forwarded to a user's pager, while messages with low priority wait until the user checks them. Horvitz' attentional interfaces are characterized by their ability to *reason* about user attention as a resource, rather than *sense* attention for a device.

7.3. Communication of attention

GAZE (Vertegaal, 1999) was one of the first Attentive Groupware Systems. Using eye trackers, it *communicates the visual attention* of the participants during mediated group collaborations. GAZE treats Awareness (Vertegaal, Velichkovsky, & Van der Veer, 1997) as an attentive phenomenon, and has been fundamental to a vision in which not just communication systems, but *all* computing systems communicate attention. In GAZE-2 (Vertegaal, Weevers, Sohn, & Cheung, 2003), streaming media optimize bandwidth according to the user's visual attention. Video images of users zoom on the basis of visual interest, and audio connections introduce an artificial Cocktail Party Effect on the basis of the visual interest of the group.

7.4. Gradual negotiation of turns

Simple user interest tracker (*SUITOR*) (Maglio, Barrett, Campbell, & Selker, 2000a, 2000b) was one of the first Attentive Information Systems. *SUITOR* provides a GUI architecture that tracks the attention of users through multiple channels, such as eye tracking, web browsing and application use. It uses this to model the possible interest of the user, so as to present her with suggestions and web links pertaining to her task. In order not to interfere with the user's foreground task, it displays all suggestions using a small ticker tape display at the bottom of the screen. SUITOR shows the importance of modeling *multiple* channels of user behavior and demonstrates how to use a peripheral low-density display to avoid interrupting a user with information of which the relevance to her foreground task is not fully known.

Pong is a robot head that rotates to face users by tracking pupils with a camera located in its nose (Morimoto et al., 2000). *FRED* (Vertegaal et al., 2001) is an Attentive Embodied Conversational System that uses multiple animated head models to represent agents on a screen. Agents track eye contact with a user to determine when to take turns. Pong & FRED show how anthropomorphic cues from head and eye activity may be used to signal device attention to a user, and how speech engines can track eye contact to distinguish what entity a user is talking to. FRED shows how proximity cues may be used to move from foreground to peripheral display with malleable resolution. When the user stops talking and fixating at an agent it looks away, and shrinks to a corner of the screen. When users produce prolonged fixations at an agent and start talking, the agent makes eye contact and moves to the foreground of the display. Maglio et al. (2000a, 2000b) and Oh et al. (2002) demonstrated that when issuing spoken commands, users do indeed look at the individual devices that execute the associated tasks. This means eye contact sensing can be used to open and close communication channels between users and remote devices, a principle known as *Look-to-Talk*. EyeR (Selker, 2001) is a pair of tracking glasses designed for this purpose. By emitting and sensing infrared beams, these glasses detect when people orient their head towards another device or user with EyeR. EyeR does not sense eye position. It stimulated us to develop eye trackers suitable for Look-to-Talk: low-cost, calibration-free, long range, wearable *eye contact sensors*.

7.5. Augmentation of attention: less is more

Attentive focus through multi-resolution vision is a fundamental property of the human eye. The acuity of our retina is highest in a 2° region around the visual axis, the *fovea*. Beyond 5°, visual acuity drops into *peripheral vision* (Duchowski, 2003). *Gaze-contingent Displays* update their images in between fixations to allow alignment of visual material with the position of the fovea, as reported by an eye tracker. Originally invented to study vision, reading and eye disease, they are now used to optimize graphics displays (Duchowski, 2003; McGonkie & Rayner, 1975). By matching the level-of-detail of 3D graphic card rendering with the resolution of the user's eye, Virtual Reality display is improved (Murphy & Duchowski, 2001). This technology inspired our design of dynamic multi-resolution windows, discussed below.

With the move towards Context-Aware Interfaces (Moran & Dourish, 2001), we are seeing increased use of attentive visualization in HCI. *Focus + Context* (Baudisch, Good, & Stewart, 2001) is a wall-sized low-resolution display with a high-resolution embedded display region. Users move graphic objects to the hi-res area for closer inspection, without loosing context provided by peripheral vision. It is an elegant example of *static* multi-resolution windows. *Popout Prism* (Suh, Woodruff, Rosenholtz, & Glass, 2002) focuses user attention on search keywords found in a document by presenting keywords throughout a document in enlarged, colored boxes. Such attentive user interfaces are distinct from context–aware interfaces in that they *focus* on designing for attention.

Architects and designers such as Mies Van Der Rohe (Carter & Mies van der Rohe, 1999), have long advocated focusing design resources in ways that provide synergies between manufacturing, human factors and aesthetic requirements. His adagio "Less is More" reflects the need to consider human attention in design. Many tools can be characterized as having been designed with attentive properties in mind. The thin blue lines that aid handwriting on paper are a good example. Since peripheral vision is least sensitive to blue detail, the lines are visible only when you need them (Duchowski, 2003). According to Goldhaber (1997), the Internet can be viewed as an economy of attention. Drawing analogies with human group communication, Goldhaber argues convincingly that buying and selling attention is its natural business model. Indeed, advertising agencies sell page views, while the *Google* search engine ranks results by the number of outside links to a page. Our framework extends upon the basic principles outlined by these designers, to create attention aware systems that truly augment the user's attention, and in so doing, his or her intellect.

8. Creating effective attentive user interfaces

As a goal, attentive user interfaces emphasize the design of interactions such that they optimize the use of the user's attentive resources. We will now describe our efforts towards the development of a number of attentive user interface prototypes, along the categorization provided.

8.1. Sensing attention: the eye contact sensor (ECS)

With the design of eye contact sensors, or ECS, we wanted to push attention sensing, in the form of eye tracking, beyond desktop use. Current desk-mounted eye trackers limit head motion of the user and do not track beyond a 60 cm distance (Duchowski, 2003), restricting the user to a localized area with little movement. In addition, head-mounted portable eye trackers are expensive, obtrusive and difficult to calibrate (Duchowski, 2003). To implement a system analogous to Look-to-Talk with ubiquitous computers, we needed a cheap ubiquitous input device that sensed *eye contact* only. The $800 eye contact sensor consists of a camera that finds pupils within its field of view using computer vision (see Fig. 3). A set of infrared LEDs is mounted around the camera lens. When flashed, these produce a *bright pupil* reflection (*red eye* effect) in eyes within range. Another set of LEDs is mounted off-axis. Flashing these produces a similar image, with black pupils. By syncing the LEDs with the camera clock, a bright and dark pupil effect is produced in alternate fields of each video frame. A simple algorithm finds any eyes in front of a user by subtracting the even and odd fields of each frame (Morimoto et al., 2000). The LEDs also produce a reflection from the surface of the eyes. These appear near the center of the detected pupils when the onlooker is looking at the camera, allowing the detection of eye contact without *any* calibration. Eye contact sensors stream information about the number and location of pupils, and whether these pupils are looking at the device over a wireless TCP/IP connection. When mounted on any ubiquitous device, the current prototype can sense eye contact with the device at up to 3 m distance. By mounting multiple eye contact sensors on a single ubiquitous device, and by networking all eye contact sensors in a room, eye fixations can be tracked with great accuracy throughout the user's environment.

8.1.1. The physiologically attentive user interface (PAUI)

Our group has also begun experimenting with specific physiological metrics which could enable us to understand the user's internal attentional state. Beyond just eye contact

Fig. 3. Eye contact sensor.

sensing, by examining such electrical signals as the electrocardiogram (ECG) from the heart, and the electroencephalogram (EEG) from the brain, we can determine specific *attentional states* that would be difficult to obtain from external data. The physiologically attentive user interface (PAUI) measures mental load using heart rate variability (HRV) signals, and motor activity using electroencephalogram (EEG) analysis (Chen & Vertegaal, 2004). The PAUI uses this information to distinguish between four attentional states of the user: at rest, moving, thinking and busy. If, for example, the user is in a busy state, then perhaps the cell phone call would not ring but merely vibrate so as not to disturb the user (see Fig. 4).

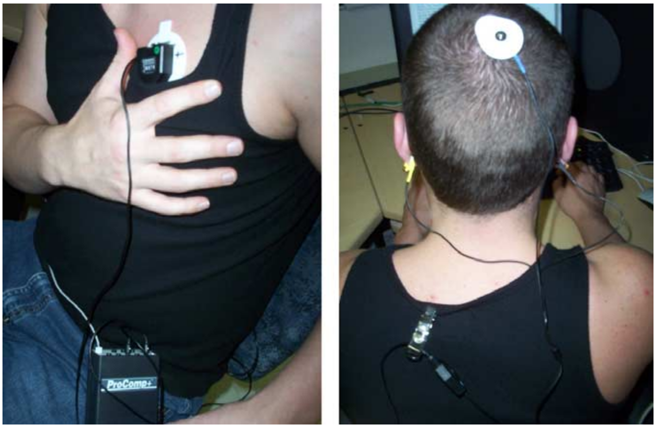


Fig. 4. PAUI heart rate monitor (left) and EEG (right).

8.2. Reasoning about turns: eyePLIANCES and eyeREASON

eyePLIANCES are smart ubiquitous appliances with embedded attention sensors, designed to extend the existing concept of gradual turn taking. Users interact with eyePLIANCES through speech, keyboard, radio tags (Want et al., 1999) or manual interaction. Functionality in appliances is accessed through X10 home automation software (X10, 2002) and wireless Internet connectivity. Fig. 5 shows the simplest form of eyePLIANCE, a light fixture appliance with an embedded eye contact sensor (Mamuji, Vertegaal, Shell, Pham, & Sohn, 2003). Using speech recognition, the light is turned on or off by saying ‘‘On’’ or ‘‘Off’’ while looking at its eye contact sensor. Using eye contact sensors as pointing devices for the real world eases problems of naming conventions for speech interaction, and juggling of remote controls (Vertegaal, Cheng, Sohn, & Mamuji, 2005). When users do use a remote or keyboard to control eyePLIANCES, eye contact sensing is used to determine the target of keyboard actions.

8.2.1. eyeREASON

Fig. 6 shows how eyePLIANCES may function in a more complex, attention-sensitive environment that keeps track of the devices users are paying attention to, the preferred notification channels, and prioritization of notifications. A personalized central server, called eyeREASON, handles all *remote* interactions of a user with devices, including user notification by devices. Devices report to the server whether a user is working with them and what that user’s focus is by tracking manual interactions and eye contact with the device. Devices may use RFID tags (Want et al., 1999) to identify and detect users and objects. Any speech I/O with a user is processed through a wireless headset by a speech recognition and production system on the server. As the user works with various devices,

Fig. 5. AuraLamp eyePLIANCE.
]

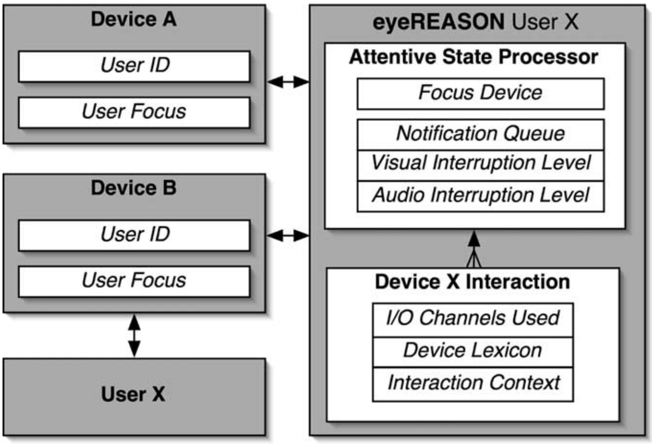


Fig. 6. eyeREASON architecture.

eyeREASON switches its context to the lexicon of the focus device, sending commands to that device's I/O channels.

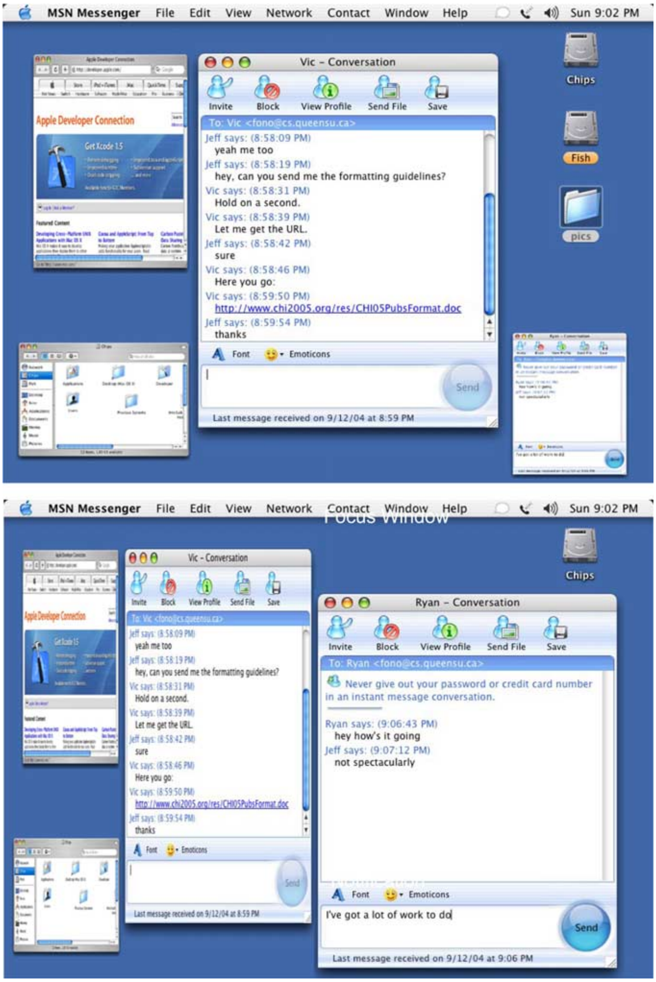
8.2.2. eyeWINDOWS

eyeWINDOWS (Fono & Vertegaal, 2004, 2005) complements the above prototypes by introducing gradual *visual* turn taking negotiated through "eye contact" with GUI windows. In eyeWINDOWS, regular windows and icons are substituted by elastic views: zooming windows of malleable resolution (Fono & Vertegaal, 2004, 2005). eyeWINDOWS automatically optimize the amount of screen real estate they use on the basis of the amount of visual attention they receive. Unlike traditional windows, the focus window is selected using eye fixations measured by a desk-mounted eye tracker. To avoid Midas Touch effects, as well as problems associated with focus targeting during magnification (Gutwin, 2002) eyeWINDOWS zoom only once the user presses an activation key.

Fig. 7 shows a desktop with eyeWINDOWS. The user is looking at the center window, which is zoomed to maximum size. Surrounding the window are thumbnails of other document windows that function as active icons. When the user looks at the bottom right window in Fig. 7, it automatically zooms to become the new focus window. Evaluations show the use of eye selection of focus windows is about twice as fast as that of hotkeys or mouse (Fono & Vertegaal, 2005).

8.3. Communicating attention

As the GAZE systems showed (Vertegaal et al., 2003), AUIs can also communicate attention to others. auraMIRROR is a media art work, a video mirror that renders the virtual windows of attention through which we interact with other people. auraMIRROR provides an ambient display which renders visualizations of virtual bubbles of attention, or *attentive auras,* that engulf groups of people during conversations, and that distinguish sub-groups in side conversations (Duncan, 1972). It unobtrusively communicates the negotiation of attention, and the effects of intrusion and interruption on this process. The mirror consists of a large plasma display mounted on a wall (see Fig. 8). This display



To: Vic <fono@cs.queensu.ca>

Jeff says: (8:58:09 PM)
yeah me too

Jeff says: (8:58:19 PM)
hey, can you send me the formatting guidelines?

Vic says: (8:58:31 PM)
Hold on a second.

Vic says: (8:58:39 PM)
Let me get the URL.

Jeff says: (8:58:42 PM)
sure

Vic says: (8:58:46 PM)
Here you go:

Vic says: (8:59:50 PM)
http://www.chi2005.org/res/CHI05PubsFormat.doc

Jeff says: (8:59:54 PM)
thanks

Last message received on 9/12/04 at 8:59 PM

To: Vic <fono@cs.queensu.ca>

Jeff says: (8:58:09 PM)
yeah me too

Jeff says: (8:58:19 PM)
hey, can you send me the formatting guidelines?

Vic says: (8:58:31 PM)
Hold on a second.

Vic says: (8:58:39 PM)
Let me get the URL.

Jeff says: (8:58:42 PM)
sure

Vic says: (8:58:46 PM)
Here you go:

Vic says: (8:59:50 PM)
http://www.chi2005.org/res/CHI05PubsFormat.doc

Jeff says: (8:59:54 PM)
thanks

Last message received on 9/12/04 at 8:59 PM

To: Ryan <fono@cs.queensu.ca>

Never give out your password or credit card number in an instant message conversation.

Ryan says: (9:06:43 PM)
hey how's it going

Jeff says: (9:07:12 PM)
not spectacularly

I've got a lot of work to do

Last message received on 9/12/04 at 9:06 PM

Fig. 7. eyeWINDOWS with zooming focus window.

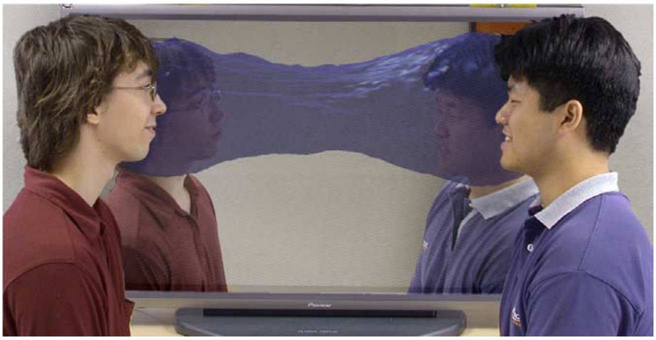


Fig. 8. auraMIRROR showing merging of auras.

reflects the world in front of it by displaying images from a video camera mounted on top of it. The images from the camera are also used to track the orientation of people standing in front of the mirror.

When two people standing in front of the mirror turn to look at each other, the virtual windows of attention between them are visualized through a merging of their auras. However, when people look at the mirror to see this, their auras break apart.

8.4. Augmenting user attention

Attentive User Interfaces can also be used to enhance the user's cognitive processes, for example, by filtering out irrelevant information before it even reaches the brain. This allows users to focus their attentive resources differently, and potentially more effectively. The Attentive Headphones, for example, are a pair of noise cancelling head phones augmented with a microphone and an ECS (see Fig. 9). Normally, noise cancelling head phones block out all noise around the user, however, this isolates the wearer from the attention of his co-workers. The eye contact sensors in the attentive headphones allow them to reason about when to turn off the noise cancellation, e.g., when somebody makes eye contact with the wearer. The Attentive Headphones create a gradual, and consequently more natural, turn taking effect in a social interaction than would otherwise be possible if auditory attention is blocked.

8.4.1. Attentive cubicles

The next step is to have *all* social interactions, including collocated ones, mediated by attention-aware systems. In office cubicle farms, where many users share the same workspace, problems of managing attention between co-workers are particularly acute. Our attentive cubicle system (Danninger, Vertegaal, Siewiorek, & Mamuji, 2005; Mamuji, Vertegaal, Dickie, Sohn, & Danninger, 2004) addresses this problem by automatically mediating auditory and visual communications between co-workers on the basis of information about their social–geometrical relationships. Our prototype cubicle's walls were constructed using a special translucent material called Privacy Glass™ (see Fig. 10). Privacy

Fig. 9. Attentive headphones.

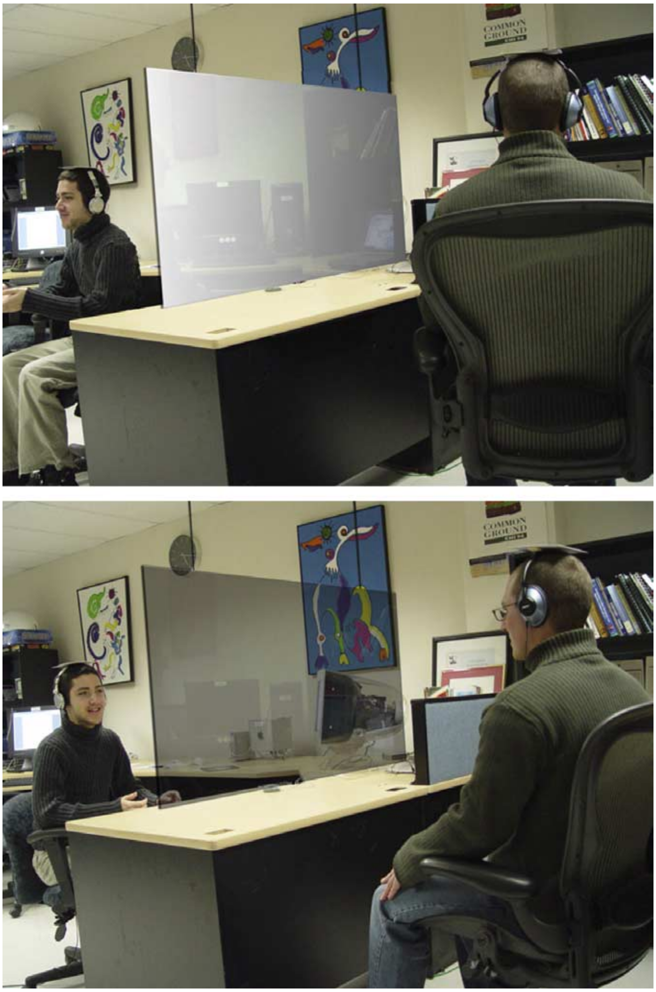


Fig. 10. Attentive office cubicle prototype in opaque (top) and transparent mode (bottom).

glass consists of a glass pane with an embedded layer of liquid crystals. When powered off, the crystals are aligned randomly, making the glass appear frosted and opaque. When a voltage is applied, the liquid crystals in the glass align, allowing light to go through the pane, thus rendering the glass transparent. When the privacy glass is opaque, cubicle workers cannot be seen by others, and are not distracted by visual stimuli from outside their cubicle. When the privacy glass is transparent, a cubicle worker can interact visually with workers on the other side of his cubicle wall. We augmented the privacy glass with a contact microphone to allow our system to detect knocks by co-workers on the pane. These knocks inform the system of a request for attention of a person inside an opaque cubicle. To mediate auditory interactions, cubicle workers wear attentive headphones. Upon detection of a request for attention, these headphones automatically become transparent to sound from the outside world.

Users within our office environment are tracked by overhead cameras mounted in the ceiling. This allows the cubicle to detect co-location and co-orientation of participants, as well as orientation towards joint objects of interest, such as whiteboards. For each tracked individual, the cubicle reports information about potential communication partners to that individual's eyeREASON server. The eyeREASON server controls the setting of the headset of the associated individual, as well as the transparency of the privacy walls of a cubicle entered by that individual.

8.4.2. Scenario

The following scenario illustrates the use of the system. User Alex is busy finishing a report. Alex has a tight deadline, as he needs to have the report filed by the end of the day. While Alex is trying to focus on his writing, his colleague Jeff is discussing a design strategy with Laurie, a co-worker, in the next cubicle. All three individuals are wearing an attentive headset that is tracked by the system. The cubicle recognizes Laurie and Jeff are co-located and oriented towards each other, without any physical barriers between them. It reports each as a potential communication partner to the other person's eyeREASON server. This causes their headphones to be set to transparent, allowing Jeff and Laurie to hear each other normally. At the same time, the cubicle detects that Alex is not co-located with anyone, and is oriented towards his computer. Alex's eyeREASON server is notified that there are no apparent communication candidates, causing it to engage noise cancellation and render his cubicle's privacy glass opaque. When Jeff and Laurie require Alex's assistance, Jeff makes a request for Alex's attention by knocking on the cubicle's privacy glass. The request is forwarded to Alex's eyeREASON server, which informs the cubicle to consider the wall between the two individuals removed. It also causes Alex's noise cancellation to be turned off temporarily, allowing him to hear the request. As Alex responds to the request, he orients himself to the source of the sound. The cubicle detects the co-orientation of Jeff and Alex. Alex's eyeREASON server renders the privacy pane between Jeff and Alex translucent, allowing them to interact normally. After the conversation is completed, Jeff moves away from the cubicle wall, continuing his discussion with Laurie. Alex turns his attention back towards his computer system, causing the cubicle to conclude Alex and Jeff are no longer candidate members of the same social group. Alex's eyeREASON server responds by turning on noise-cancellation in Alex's headset, and by rendering the privacy glass of his cubicle opaque again.

The above scenario illustrates how entire rooms can be designed to balance social as well as privacy needs of co-workers in a dynamical fashion. The above scenario can also be applied to remote situations.

8.4.3. OverHear

OverHear is a remote surveillance interface that aims to augment the user's remote auditory attention. It consists of an eye-tracking display showing a live audio and video feed obtained from a robotic directional microphone and webcam at a remote public location (see Fig. 11). When the user looks at a particular individual in the video stream, the directional microphone on the other end in the remote location will focus upon that person, allowing the user to hear that specific conversation. The OverHear interface simulates and enhances the natural cocktail party phenomenon by blocking out peripheral noise,

Fig. 11. OverHear eye tracking surveillance display (top) and robotic shotgun microphone (below).

creating a focus that augments the user's auditory attention in ways otherwise not possible.

9. Discussion

Throughout the process of designing attentive user interfaces, we came across many issues that have helped us identify outstanding research questions. Among the concepts we explored, we found the metaphor of virtual windows of attention particularly inspiring. Whether in visual or auditory interactions with remote devices or people, users need to be supported by subtle cues that make up the *virtual windows* through which entities communicate with them. It is not sufficient to define such windows by the electronic channels through which interactions take place, because electronic channels do not delineate actual attention. By sensing user attention, devices may know when users are attending to them. By providing devices with a means of communicating their attention, users may know they are being attended to as well (Bellotti et al., 2002). This allows users and devices to establish the negotiation of joint interest that is characteristic of multiparty human turn taking. We wish to invite researchers and designers to further develop and improve upon the conceptual framework provided in this paper. One of the technical problems we encountered is that of sensing attention for small or hidden devices. While physiological sensing technologies may address these issues, they are potentially invasive. A second issue is the identification of users at a distance. While eye contact sensors may one day be able to perform iris scanning, there are privacy implications that must be considered. A third is that of prioritization of notifications. Can we trust automated services to rank and prioritize the information we receive? We believe the most pressing issue relating to the sensing technologies presented in this paper is that of *privacy*. How do we safeguard privacy of the user when devices routinely sense, store and relay information about their identity, location, activities, and communications with other people?

10. Conclusions

This paper presented a framework for designing Attentive User Interfaces, user interfaces that augment the user's attention. AUIs achieve this by negotiating interactions in ubiquitous environments, where demands on our attention may exceed our capacity. By treating user attention as a limited resource, such interfaces reduce disruptive patterns of interruption. By embedding ubiquitous devices with attention sensors (such as eye contact sensors) that allow them to prioritize and manage their demands on user attention, users and devices may enter a turn taking process similar to that of human group conversation. By designing virtual windows of attention between devices and users, communications in multiparty HCI may become more sociable as well as more efficient.

References

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. London: Cambridge University Press.

Baudisch, P., Good, N. & Stewart, P. (2001). Focus plus context screens: Combining display technology with visualization techniques. In *Proceedings of UIST'01*, Orlando.

Bellotti, V. et al. (2002). Making sense of sensing systems: Five questions for designers and researchers. In *Proceedings of CHI 2002* (pp. 415–422). Minneapolis: ACM Press.

Bolt, R. A. (1985). Conversing with computers. *Technology Review, 88*(2).

Carter, P., & Mies van der Rohe, L. (1999). *Mies van der rohe at work*. Phaidon Press.

Cherry, C. (1953). Some experiments on the reception of speech with one and with two ears. *Journal of the Acoustic Society of America, 25*, 975–979.

Chen, D., & Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive user interfaces. In *Proceedings of CHI 2004* (pp. 1513–1516). Vienna: ACM Press.

Danninger, M., Vertegaal, R., Siewiorek, D., & Mamuji, A. (2005). Using social geometry to manage interruptions and co-worker attention in office environments. In *Proceedings of Graphics Interface 2005*. Victoria, Canada.

Duchowski, A. (2003). *Eye tracking methodology: Theory and practice*. Berlin: Springer-Verlag.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personal and Social Psychology*, 23.

Fono, D., & Vertegaal, R. (2004). EyeWindows: Using eye-controlled zooming windows for focus selection. In *Proceedings of UIST 2004. In Video proceedings*. Santa Fe: ACM Press.

Fono, D., & Vertegaal, R. (2005). EyeWindows: Evaluation of eye-controlled zooming windows for focus selection. In *Proceedings of CHI 2005*. Portland: ACM Press.

Gibbs, W. (2005). Considerate computing. *Scientific American, 292*(1), 54–61.

Goldhaber, M. (1997). The attention economy and the net. First Monday. Available from: <http://www.firstmonday.dk>.

Gutwin, C. (2002). Improving focus targeting in interactive fisheye views. In *Proceedings of CHI 2002* (pp. 267–274). Minneapolis: ACM Press.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In *Proceedings of CHI'99*. Pittsburgh: ACM Press.

Horvitz, E., Jacobs, A., & Hovel, D. (1999). Attention-sensitive alerting. In *Proceedings of UAI'99* (pp. 305–313). Stockholm: Morgan Kaufmann.

Ishii, H., & Ullmer, B. (1997). Tangible Bits: Towards seamless interfaces between people, bits, and atoms. In *Proceedings of CHI'97* (pp. 234–241). Atlanta: ACM Press.

Jacob, R. (1995). Eye tracking in advanced interface design. In W. Barfield & T. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 258–288). New York: Oxford University Press.

Maglio, P., Barrett, R., Campbell, C., & Selker, T. (2000). SUITOR: An attentive information system. In *Proceedings of the international conference on intelligent user interfaces*.

Maglio, P., Matlock, T., et al. (2000b). Gaze and speech in attentive user interfaces. In *Proceedings of the international conference on multimodal interfaces*. Berlin: Springer-Verlag.

Mamuji, A., Vertegaal, R., Shell, J., Pham, T., & Sohn, C. (2003). AuraLamp: Contextual speech recognition in an eye contact sensing light appliance. In *Extended abstracts of ubicomp 03*.

Mamuji, A., Vertegaal, R., Dickie, C., Sohn, C., & Danninger, M. (2004). Attentive office cubicles: Mediating visual and auditory interactions between office co-workers. In *Video proceedings of ubicomp 2004*.

McGonkie, G. W., & Rayner, K. (1975). The span of effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578–586.

Moran, T. P., & Dourish, P. (Eds.)(2001). Context-aware computing [Special issue] *Human Computer Interaction, 16*.

Morimoto, C., et al. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing, 18*.

Murphy, H., & Duchowski, A. (2001). Gaze-contingent level of detail rendering. In *Proceedings of the eurographics 2001*. Manchester.

Nielsen, J. (1993). Noncommand user interfaces. *Communications of ACM, 36*(4), 83–99.

Oh, A. et al. (2002). Evaluating look-to-talk: A gaze-aware interface in a collaborative environment. In *Extended abstracts of CHI 2002* (pp. 650–651). Seattle: ACM.

Selker, T. (2001). Eye-R, a glasses-mounted eye motion detection interface. In *Extended abstracts of CHI 2001*. Seattle: ACM Press.

Shell, J. S., Selker, T., & Vertegaal, R. (2003). Interacting with groups of computers. *Communications of the ACM, 46*(43), 40–46.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London.

Smith, D., Irby, C., et al. (1982). Designing the star user interface. *BYTE, 7*(4).

Suh, B., Woodruff, A., Rosenholtz, R., & Glass, A. (2002). Popout prism: Adding perceptual principles to overview + detail document interfaces. In *Proceedings of CHI 2002* (pp. 251–258). Minneapolis: ACM Press.

Vertegaal, R., & Ding, Y. (2002). Explaining effects of eye gaze on mediated group conversations: Amount or synchronization? In *Proceedings of CSCW 2002*. New Orleans: ACM Press.

Vertegaal, R. (1999). The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of CHI'99*. Pittsburg: ACM Press.

Vertegaal, R., Slagter, R., Van der Veer, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In *Proceedings of CHI 2001* (pp. 301–308). Seattle: ACM Press.

Vertegaal, R., Velichkovsky, B., & Van der Veer, G. (1997). Catching the eye: management of joint attention in cooperative work. *SIGCHI Bulletin, 29*(4).

Vertegaal, R., Weevers, I., Sohn, C., & Cheung, C. (2003). GAZE-2: Conveying eye contact in group videoconferencing using eye-controlled camera direction. In *Proceedings of CHI 2003*. Fort Lauderdale, FL: ACM Press.

Vertegaal, R., Cheng, D., Sohn, C., & Mamuji, A. (2005). Media eyepliances: Using eye tracking for remote control focus selection of appliances. In *Extended abstract of ACM CHI 2005*. Portland, OR: ACM Press.

Want, R., Fishkin, K., et al. (1999). Bridging physical and virtual worlds with electronic tags. In *Proceedings of CHI'99* (pp. 370–377). Pittsburgh: ACM Press.

Weiser, M. (1991). The computer for the 21st century. *Scientific American, 265*(3), 94–104.

X10. (2002). *Home automation*. Available from: <http://www.x10.com>.

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of CHI'99* (pp. 246–253). Pittsburgh: ACM Press.