A Framework for Supporting Joint Interpersonal Attention in Distributed Groups

Jeremy Birnholtz, Johnathon Schultz, Matthew Lepage, Carl Gutwin

A Framework for Supporting Joint Interpersonal Attention in Distributed Groups

Jeremy Birnholtz, Johnathon Schultz, Matthew Lepage and Carl Gutwin

Departments of Communication and Information Science
Cornell University, 336 Kennedy Hall, Ithaca NY, USA

Department of Computer Science, University of Saskatchewan
110 Science Place, Saskatoon, SK, S7N 5C9, Canada
jpb277, jts228, mcl82 @ cornell.edu; carl.gutwin@usask.ca

Abstract. Informal interactions are a key element of workgroup communication, but have proven difficult to support in distributed groups. One reason for this is that existing systems have focused either on novel means for gathering information about the availability or activity of others, or on allowing people to display their activities to others. There has not been sufficient focus on the interplay between these activities. This interplay is important, however, because mutual awareness and attention are the mechanisms by which people negotiate the start of conversations. In this paper, we present the OpenMessenger Framework, a system and design framework rooted in the assumption that individual behaviors occur in anticipation of and/or in response to the behavior of others. We describe both the system architecture, and specific examples of the novel implementations it enables. These include techniques for coupling gathering behaviors with display behaviors, and for integrating these into user workspaces via peripheral displays and gaze tracking.

Keywords: Awareness, Attention, Interaction, CMC, CSCW.

Introduction

Informal interaction has repeatedly been shown by CSCW researchers to be a key attribute of modern work [2, 37, 47]. Significant efforts over the past 20 years have focused on supporting these interactions via improved awareness of others' presence and activities [21, 31], by improving people's ability to interrupt at appropriate times [19], and by strategically displaying impending interruptions (e.g., [33, 38]). While this work has yielded many research prototypes, the most common tools used in everyday, real-world collaboration still offer only rudimentary support for initiating and concluding informal interactions [9, 46].

One key reason for this is the persistent difficulty of supporting fluid transitions between passive awareness of one's surroundings and engaged attention to a particular person or object. Many have demonstrated the ability of people in face-to-face environments to track others' activities in their visual or auditory periphery, and adjust their own activities or shift their focus of attention accordingly (e.g., [28, 45]).

These behaviors have proven difficult to support online, however [31, 46]. Nearly ten years ago, Schmidt [44] discussed the problems of framing awareness either as an abstract sense of others that is independent of attentional focus *or* as the object of attention itself. He noted that people are constantly monitoring their environment, but also sometimes focusing on others and adjusting behavior accordingly. Since Schmidt's discussion of these issues, however, few systems or frameworks have addressed this dual nature of awareness.

We argue that one way to address this issue is by considering it as a problem of joint attention management, drawing on Clark's [17] notion of *joint activity.* Joint activity occurs when two or more people act individually but coordinate their actions toward achievement of a shared goal. In assessing availability and initiating conversation, we can think of individual actions in terms of *gathering* (Schmidt [44] calls this "monitoring") information about others' availability and signaling, or *displaying* availability for or interest in interaction.

Suppose Alex and Bill work across the room from each other. If Alex moves closer to Bill to *gather* information about Bill's availability, Bill may notice Alex's presence and glance at him. Alex then notices Bill glance and returns it, or, if Bill seems busy, Alex may decide to come back another time. In this way, Alex's approach simultaneously serves as *gathering* and *display*. Closer proximity means that Alex can gather more information, and makes Alex's presence more noticeable to Bill. This triggers Bill's glance, which allows Bill to see that Alex is approaching, and to *display* by glancing at Alex that Bill has noticed Alex's approach.

In contrast to this joint perspective, most existing systems have focused *either* on gathering (e.g., deciding when to interrupt [19, 40]) *or* display (e.g., notifying others of one's status [14, 38]). This separation makes it hard to understand or support the interactive aspects of paying attention via acts of gathering and display. In this paper, we introduce the OpenMessenger Framework (OMF), a flexible and extensible framework for developing joint-attention systems. Through OMF, we define and implement structures to address three problems: 1) discerning the user's focus of attention, and treating this differently when focus is on another person; 2) allowing for easy joint action both during and prior to conversational interaction; and 3) allowing for easy and natural awareness of other users' presence and behavior.

Background

When people are aware of or interact with others, we consider them to be managing their attention, which can vary both in intensity and focus, to activities and stimuli in the environment. As such, we define attention more broadly than a purely cognitivist interpretation that considers only a single point of focus [35]. Broadening this definition allows us to account for transitions between more and less active attentional states, and is consistent with evidence from neuroscience on the mechanisms by which people sense and attend to the presence of others (e.g., [41]).

At the same time, our treatment leaves aside debate on the relationship between attention and awareness, as discussed by Schmidt [44]. As Schmidt ultimately points out, however, the interesting practical questions in this domain are not about the

underlying awareness and attention mechanisms, but rather the roles played by specific types of information in affecting behavior. We take as our baseline assumption that abstract awareness in the sense discussed by Dourish and Bly [22] often leads to conscious and focused attention, and that the details of this transition from awareness to engagement are not well supported by today's tools.

We look at these processes within the context of informal interaction and conversation initiation. This is not an exhaustive treatment, but represents a first order approximation [1] of a social problem; one that will enable a foothold in this key area.

Gathering, Display and Coupling

Attentional behaviors can be usefully framed using the terms *gathering* and *display*. *Gathering* refers to one person getting (either via passive monitoring or active seeking) information about what another is attending to, such as tasks, other people or the gatherer herself. *Display* refers to information about the focus of one's attention that is available for others to gather. We separate these concepts for discussion, but note that they are highly interdependent.

These cues are used reciprocally; initiating interaction is a process of negotiating joint attention via sequences of actions informed by present and prior actions of others [10, 28, 36]. As such, one framework that can be useful in understanding this process is Clark's framework of joint action. This framework allows for the situation of Schmidt's observations about both passive and active states of awareness within the context of behavior aimed at a shared objective or goal. Joint action occurs when people act in the belief that they are part of a collective activity, in which their actions occur in response to coordination signals from another party. We argue that gathering and display are what Clark [17] calls component moves in joint action.

Consider again a case in which Alex wants to talk to Bill. Alex must get Bill's attention, which he does using both *gathering* and *display*. Glancing quickly allows Alex to *gather* information about Bill's likely availability; he then walks toward Bill, which serves to *display* Alex's intent to start a conversation. Bill then can respond by using gaze and body position to *display* his own interest.

These individual actions can be interpreted within the framework of joint activity. If Bill is wearing headphones and does not look up as Alex approaches, for example, that could signal either that Bill is unaware of Alex, meaning that Alex needs to get closer and louder; or that Bill is aware that Alex is trying to talk to him, but signaling that he (Bill) is unavailable. This judgment on Alex's part is based on information about Bill, the outcomes of recent actions, and knowledge of the context.

One key attribute of Alex's approach for the purposes of attention management is the relationship between gathering and display. Alex's movement toward Bill to gather information necessarily functions simultaneously as an act of display, because Bill can see him approaching [10]. We can therefore say that these instances of gathering and display are tightly coupled. In face-to-face interactions, people's ability to notice others (e.g., [41]) relies on this coupling. We perceive others' gathering because it is visible in ways that we can attend to.

Returning to Clark's terms, we can define coupling as the extent to which a particular action is visible and noticeable to others. Table 1 illustrates coupling

relationships between gathering and display. In face-to-face approaches (top left cell), physical proximity and eye gaze are effective ways to display attention because gathering and display are tightly coupled [10]. In contrast, many behaviors that are tightly coupled in face-to-face interactions have different relationships online [5], as illustrated in the three remaining cells of Table 1. If Alex and Bill are spying covertly on each other via webcam, for example, gathering and display may be completely de-coupled (bottom right cell). Alex's gathering is not displayed at all, and vice versa.

		Alex Gathers	
		Displayed	**Not Displayed**
Bill Gathers	**Displayed**	Face-to-Face Approach	Some IM conversations
	Not Displayed	Some IM conversations	Spying, covert looking

Table 1. Coupling of Gathering and Display

This helps to address a key question identified by Schmidt [44]: how do people regulate the obtrusiveness of their behavior? In face-to-face encounters, obtrusiveness is often a function of the salience of display. Staring at somebody's screen or standing very close to them, for example, are very salient forms of display that are also obtrusive [10]. Online, however, it is possible to get detailed information about others without appearing obtrusive or even being visible at all. These examples of asymmetric coupling relationships have a significant effect both on how people regulate their behavior, and on the feasibility of joint action.

One example is the "appear offline" option on instant messaging (IM) clients used by those wishing to avoid interruptions [9]. People using this option can gather information about others on their contact list without those others even knowing that such gathering is possible (because they cannot see that the gatherer is online). This sidesteps the key role that obtrusiveness ordinarily plays in attracting and negotiating attention and has significant consequences for joint action, because people cannot respond to actions of which they are unaware. In other words, it is the coupling of gathering and display in face-to-face interactions that helps facilitate joint action.

Gathering is Display; Display is Gathering

Despite Schmidt's discussion of awareness as a duality of gathering *and* display, most systems and theoretical frameworks for addressing the problem have focused either on one or the other.

Early media space systems used cameras to provide video views of others in their offices [22, 23, 29, 31]. Cameras, however, were thought by some to be invasive [15, 18]; and the systems did not support the subtleties of negotiating interaction [32]. These problems reflect the de-coupling of gathering and display in that one user could view (i.e., gather) video of another, without a clear display that this was taking place.

These early video experiences led many to experiment with the notion of a "virtual approach" (reviewed in [46]). In our terms, this work can be characterized as an attempt to increase coupling between gathering and display by displaying to an observed party that gathering is taking place, and that conversation may be desired. The idea was that the approach would facilitate interaction more naturally by allowing for multiple levels of gathering, and by displaying activities to the observed parties.

As such, “approaches” often involved replicating the sequence of actions typically involved in initiating face-to-face conversations. Several systems allowed users to, for example, “glance” at others to discern their availability [34], and then follow a series of progressively more informative steps eventually resulting in conversation. At each step, the observed party would have to respond in kind (e.g., with a “glance” of their own) to proceed with the interaction.

More recent systems such as Community Bar [39] allow for a continuum of awareness states. In Community Bar, these must be manually updated via independent “focus” (how much information is seen about others) and “nimbus” (how much information is revealed to others) sliders. These terms come from work by Benford and Fahlen [8] and Rodden [43] that aims to distinguish between being the object of somebody’s attention (their nimbus) and focusing on somebody (one’s focus). While this distinction is a useful one, the Community Bar implementation is problematic in that it renders the coupling relationship between gathering and display dependent on the combined status of independent users’ sliders being manipulated in parallel. That is, Alex could choose to reveal more information to Bill (via a nimbus slider) even as Bill is reducing the amount of information he sees about Alex (via a focus slider).

From a joint action standpoint, designs that foster multiple levels of awareness and sharing are an improvement in that they allow for coordinated activity. These systems were critiqued, however, for requiring lockstep and seemingly artificial sequences of behavior. By this, we mean that real-world gathering behaviors (e.g., glancing, walking up to somebody) are easily noticed and responded to because, unlike pointing at buttons in an interface or watching a PC window for notification of an incoming virtual “glance,” they are the actions that naturally occur to assess availability, respond to somebody to avoid appearing rude, and/or to start a conversation [28, 36].

We argue that effective support for a joint action approach to attention requires consideration of the interplay between gathering and display. Specifically: 1) acts of gathering must be coupled to displays or notifications, and 2) these displays must be easily noticed and responded to via subsequent acts of gathering, that must also be displayed. Gathering must be displayed and displays must be gatherable, *ad infinitum*.

Approaching is Interacting

A second key question in this area concerns the specific information that causes people to adapt their behavior in response to the actions or reactions of others. There are useful lessons from face to face interaction that can be considered here.

Goffman [30] argues that human behavior around others is performative; it is often intended to convey information or impressions to others. Sudnow [45] discusses the importance of glances in assessing others’ behavior and availability. He notes that people in public settings know that others may glance at them and act accordingly, such as by putting on headphones or adjusting posture to appear busy [10]. Sudnow [45] refers to these as “glanceable states,” in that status can be discerned via a glance.

In some ways, the glanceable state is reflected in current interaction tools, such as the IM contact list. The intent of the list is to concisely summarize who is online and available. There has also been work aimed at improving this information by automatically updating status information (i.e., availability) via sensors (e.g., [7, 25]).

One problem with the IM contact list, however, is that recent work has focused *either* on the problem of interruptions [40] and developing systems that allow for better timing of interruptions (i.e., better gathering of information; [4, 19]) *or* on techniques for unobtrusive notification of impending interruption (i.e., better display; [6, 14, 38]. Considering these problems in isolation ignores a key component of our joint action argument: acts of gathering and display occur in response to each other. Approaching somebody does not occur prior to interacting; it is part of the joint action of initiating conversation. That is, approaching *is* interacting.

People do not respond the same way to interruptions from all others, such as work collaborators vs. social friends, [20], and also may behave differently when they know their behavior is being monitored by others [27]. Moreover, not all interactions result from interruptions; many result from serendipitous mutual attention [36]. Thus, our second argument is that each act of gathering and display, however preliminary from the standpoint of starting verbal conversation, must be considered as component behaviors in an interaction (or joint action) to which others should be able to respond.

A real-world approach progresses from distant observation – characterized by less detailed gathering and less salient display – to closer observation – characterized by more detailed gathering and more salient display due to physical closeness. We advocate a similar progression online, consisting of multiple types of interactive behavior that enable both gathering and display to take place. As with the face-to-face approach, we emphasize that what is important is not reciprocal instances of identical behavior (i.e., a glance must be followed by a glance), but rather a general correlation structure between behaviors such that the amount of detail that can be gleaned from a particular gathering behavior roughly correlates with the salience of the display behavior with which that gathering is coupled.

The OpenMessenger Framework

Supporting a joint-action approach to attention management requires mechanisms for coupling gathering and display behaviors, and for treating these as interaction. In this section of the paper, we present the OpenMessenger[1] Framework (OMF), a software framework and application for addressing these issues. We aim to make two contributions: 1) an extensible software framework for experimental exploration of issues related to awareness, and 2) an implementation example with novel gathering and display mechanisms.

Supporting Gathering and Display at the Framework Level

To effectively support joint action in attention management, we need a conceptual architecture that supports gathering and display of information, and the coupling of these behaviors to each other. This is accomplished in OMF with abstractions called *sensor managers, monitors*, *awareness events,* and *views*. Data about user activities is

[1] Note that the word "open" in OMF refers to open-plan offices, which were our inspiration for this work. We will happily share OMF source code, but "open" does not imply open-source.

captured from hardware sensors by *sensor managers,* and is then analyzed and distributed to other OM clients by *monitors*. Transmission of the data is via *awareness events* passed from clients to the OM server, and then to all clients in a pre-defined group (see Figure 1). Event information is made perceptually available to users in a *view* that could be a visual, auditory, or tactile display.

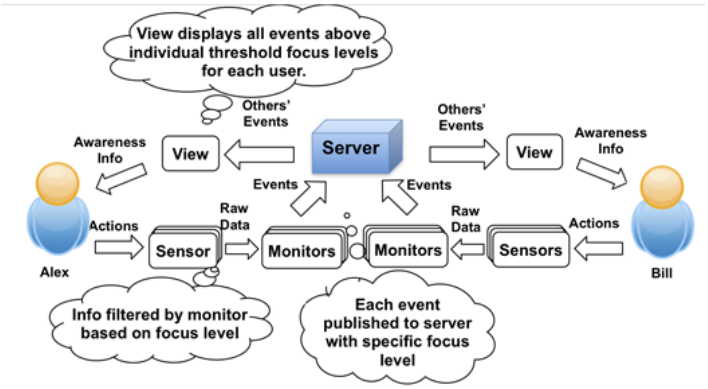

Figure 1. Information flow in the OpenMessenger Framework

Sensor Managers

Gathering information about others is supported by *sensor managers,* which are C# objects on the client PC that regularly sample data sources in the environment such as hardware devices or operating system information. Sensor managers provide information to other components using a publish/subscribe model. Subtypes of the SensorManager class are singletons.

Monitors

To allow people to gather useful information about others while still restricting the flow of potentially sensitive sensor data (as in [7, 24]), and avoiding user inundation with information, OMF uses *monitors* to distill raw sensor data. OMF monitors are C# objects on the client that subscribe to one or more sensor managers and process the raw sensor data for publishing to the server as *awareness events*. These events are sent to the server, which broadcasts them to all connected clients (including the sender, though the sender does not use them). Windows Communication Framework (WCF) is used for network communication (based on TCP sockets). When an awareness event is received, the server stores it in the server *event cache*. With this cache, event data can be immediately sent to newly connected clients.

As an example, the monitor for a keyboard activity sensor might release keystroke frequency, but not reveal which keys were pressed. This allows for conveying information about activities without releasing potentially sensitive data, and also takes data that may not be meaningful in raw form (e.g., sound level) and processes it to provide useful information (e.g., presence of sound above a conversational threshold).

Importantly, monitors are abstractions of sensor managers. While raw sensor data generally comes only in one form, data can be used in many ways. For example, a microphone sensor could indicate sound above a threshold, while another monitor could use speech recognition on the same data to determine if there is a conversation going on. As with sensor managers, subtypes of the Monitor class are singletons.

Social Monitors

Monitors, as described so far, operate similarly to previous systems when the user is focused on objects in the task or environment [26]. Our joint-action perspective, however, requires a way for users to dynamically detect and respond to attention from other users. Thus, a unique feature of OMF is special support for situations when the user's focus of attention is on another user. This is accomplished via *social monitors*, a sub-type of the monitor class that communicates this information.

Social monitors are necessary because attention to another person is distinct from attention to a task element or interface object. As discussed above, attention to people is interactive; it often occurs often in expectation of and/or in response to another act of gathering or display. From an implementation standpoint, social monitors are unique in that they affect the intensity of attention that one user is paying to another, which we define later as *focus level*. Social monitors determine when one user is attending (e.g., via mouse or other input data) to another user's representation (e.g., an on-screen avatar) on a view. This is accomplished via data about attentional focus (e.g., an x,y coordinate pair from the mouse or eye tracker monitor) and knowledge about the arrangement of avatars on the view. When one user is determined to be attending to another, focus levels are adjusted accordingly (see below).

Monitor	**SensorManager Used (API)**	**Analysis**
• Typing Activity • Keypresses	Keyboard (DirectInput)	• # keys pressed since last event • Record keys since last event
• Screen Activity • Current Window • Screen Contents	Screen (Windows API)	• Compare successive screenshot frames • Record title of current focus window • Record screen image
• Sound Presence • Sound Level • Number of Speakers	Microphone (DirectSound)	• Compare microphone level to thresholds • Track level over time • Analyze sound for voices
• Overall Presence	Keyboard, Screen, Microphone	• Time since last sensor change
• Visual Presence • Visual Changes • Number of People	Webcam (DirectX)	• Compare successive frames for changes • Analyze frame for objects (e.g., people) • Record webcam image
• Mouse Focus Target	Mouse (UI toolkit)	• Determine avatar mouse is pointing at
• Eye Focus Target	Eye Tracker (Custom API)	• Determine avatar the user is looking at
• In Conversation	Chat Window (OMF events)	• Determine who the user is conversing with

Table 2: Example monitors and associated sensor managers in OMF.

Sensor and Monitor Examples

Example OMF sensor managers and monitors are shown in Table 1. This is not an exhaustive list of possibilities; most of these specific examples were chosen based on our own experience with the first OM application and based on prior work on predicting availability with sensor data [4, 7, 26]. That work suggests that keyboard activity, idle time, and the presence of sound can be very helpful in determining availability. Gathering active window titles and screen snapshots are sources that were used in the initial version of OM, but were seen by some users as too invasive [11]. We therefore added a monitor for this data source that only detects screen changes, rather than transmitting the actual screen image. The eye tracker and mouse monitors are unique to the OMF in that they are social monitors, described below.

OMF also provides a simple extension mechanism for adding new monitors for additional analyses, or new sensor managers for new data sources. This requires some coding by the application programmer – creating a new subtype of Monitor or SensorManager, and linking APIs for new hardware sensors to these classes – but the process is simple and existing classes can be used as templates.

Views

Gathering and display behaviors are rooted in the perceptual availability of awareness information. Gathering cannot occur if this information cannot be perceived, and display has no purpose if the information is unavailable to others. Awareness information in OMF is made perceptually available via a *view*. OMF provides a basic visual view based on earlier OM systems [11]. However, application programmers can easily develop new views within the .NET environment (e.g., using XAML or Windows Forms), or can build displays that use sound or other feedback mechanisms. Views are client-specific (i.e., they are not WYSIWIS), and multiple views can run simultaneously on the same client. A standard data structure for each user simplifies the storage of view-based information such as an avatar image, user name and ID, and the most recent awareness event data received pertaining to that user.

Focus Level: Degree of Attention and Coupling

As noted above, people do not respond the same way to interruptions from different people, and some interruptions result from serendipitous mutual attention. Alex may be available to Bill but not to Cathy, for example. Support for variation in treatment of contacts, and for awareness of mutual attention requires the capacity to share different information with certain contacts. Moreover, as Alex receives attention from Bill, he should be able to easily respond by displaying his own attention to Bill (indicating interest or availability) or to a task or another user (indicating that he is busy). In OMF, this is accomplished via the *focus level* mechanism, which represents the level of attention that one user is paying to another.

It is through focus levels that acts of gathering are coupled to acts of display, and that the correlation structure of gathering and display behaviors is maintained. First, the focus level of one user on another determines what information the first can gather about the second. Each awareness event includes a threshold focus level for event

data display. This means that if Alex releases an awareness event with threshold level X, Bill will only see this event if his focus level on Alex is greater than or equal to X. More detailed information is therefore assigned a higher focus level.

Second, the focus level of one user on another is used to notify the observed user (i.e., display to them) that gathering is taking place. Notification increases in salience as focus level increases. That is, Alex receives increasingly salient notifications as Bill's focus on him increases. Our OMNI system described below uses visual salience, but this could be accomplished with sounds or other cues as well. The key is that more detailed gathering behaviors correlate with more salient displays.

Focus level is represented in the system as an integer (range: 0-5, with 0 as minimum) for each user's level of focus on each other user. Focus level is updated dynamically via social monitors, which detect one user's focus on another. The 0-5 range is based on experience with the initial OM version, suggesting that six levels provides enough variance for multiple means of gathering, and for displaying varying levels of interest without distracting the user. Further, focus is uni-directional. That is, Bill's focus on Alex does not vary directly with Alex's focus on Bill. Rather, each client stores a list of the current client's focus level on all other users. As such there are two discrete focus level variables that describe Alex and Bill's focus on each other. Thus, OMF is not a strict reciprocity-based system [46].

Information displayed at each focus level in the current OM is listed in Table 3. This sequence is preliminary, but based on our experience with sharing similar information in previous OM versions. The focus level mechanism could also be exploited in more powerful ways by providing multiple information streams at each level, or by providing progressively more detailed information from a single sensor.

Focus Level	Data/Monitor (see Table 1)
0	Overall Presence
1	Keyboard Activity, Idle Time
2	Screen Activity
3	Current Window Title
4	Microphone Activity
5	Current Gaze Focus

Table 3. Example focus levels and corresponding monitors. Note that information display is cumulative, such that each level includes the lower levels as well.

Changing Focus Level

Focus level is increased via explicit indicators of attention from social monitors (e.g., mouse and eye-gaze data) and decreased over time via attenuation. With the mouse social monitor, for example, if Alex's cursor hovers over Bill's avatar for a predetermined length of time the social monitor produces a *focus change event* that is distributed to all clients.

Testing of earlier OM versions [11, 13] suggests that these trigger mechanisms must be carefully designed. For example, users tended not to use OM's awareness features if it took more than a second for information to appear. We therefore increase focus level after one second of sustained attention to an avatar, and subsequently increment focus level by one after each second of continued attention. This continues until the observer stops focusing on the target, or focus level reaches its maximum.

Focus then decreases over time at a linear rate when the observer stops focusing on the target. Decrease is slower than increase: 15 seconds to decrease from level 5 to 0, compared with 5 seconds to increase from level 0 to 5.

As with views, the mechanisms for changing focus levels can be extended in OMF. For example, it would be simple for an application programmer to add a social monitor that changes focus levels through explicit keystrokes or commands, or that changed the way that focus level increases and decreases.

Event Filtering

As noted above, monitors generate events to be displayed only to users with at least a threshold focus level on the user generating the event. This presents the problem of how to broadcast events: should they be shared with all users, but displayed only to those viewing at the appropriate focus level (receiver-filter), or shared only with users at a particular focus level (sender-filter)? We chose the receiver-filter strategy to reduce the complexity of the system in terms of server query overhead. Given relatively low numbers of connected clients and sensors, broadcasting the data to all clients at once involves fewer operations and is simpler than a query-based model. This might be reconsidered, however, in an environment with a substantially greater number of clients or sensors. This means more network traffic, but this is acceptable since overall bandwidth requirements are low.

Text Chat

Commentaire [J1]: Not sure this should be deleted.

To support conversation, the OMF provides basic text chat. This is supported by the framework via the MessageEvent, a type of client-generated OM event characterized by a short text message, a sender and a receiver. The chat window implementation is similar to other text-based chatting systems, so we do not provide a detailed description here. However, OMF chat is novel in that conversations are a potential data source – that is, the 'In Conversation' monitor keeps track of who is talking to whom (Table 1), and can use this to help determine attention and focus level.

An OMF Implementation Example: Eye Tracking and OMNI

To illustrate the capabilities of the OMF, we here describe our current implementation of a system called OMNI, which combines a social monitor that reads data from a head-mounted eye tracker, and a projected peripheral-vision awareness view (Figure 3). OMNI is an OMF implementation example intended for information or other office workers who work primarily at a desk but benefit from frequent informal interaction with potentially remote colleagues (e.g., designers [10], engineers, or researchers [2, 37]). Those with a different work environment may benefit from a different OMF implementation. Elements of this system have been reported previously [12], but this is the first explication of OMNI at the system level.

Figure 2. Screen shots of two OMNI displays in use, with rows denoting what users Alex and Bill see. In Alex's view, as Alex focuses on Bill, a halo appears around Bill's avatar along with a line to indicate that he is being focused on. On Bill's screen Alex moves closer to Bill's avatar as Alex's focus level on Bill increases. (A) is the remote user's avatar, (B) is the local user, and (C) is sensor information about the remote user.

The EyeTracker Social Monitor

One problem that the OMF enables us to address is the difficulty of mapping natural face-to-face behaviors (e.g., glancing) to online interfaces. Joint action requires the capacity for easy response, and one way we accomplish this is by using an eye/head tracker to capture the user's visual focus. Eye tracking is particularly appropriate for this context because interpersonal attention is often conveyed via the eyes. Looking at somebody implies interest, and looking away implies disinterest [3, 28, 36].

We use an ASL H6 head-mounted eye tracker and a Flock of Birds magnetic head tracker in our current system. We acknowledge this is a cumbersome and expensive device, but note that eye tracking technology is rapidly falling in cost as it becomes possible to track unobtrusively and inexpensively via webcams (e.g., [16]). While our specific implementation necessarily relies on the details of our hardware, the general approach to eye tracking that we describe here is not device-dependent; it could be adapted to any device that provides (x,y) coordinate pairs on a defined plane in space.

The eye tracker's API libraries allow for the definition of planes in the real-world space. Each eye-tracker data point consists an (x,y) coordinate pair on a particular plane. Augmented by basic knowledge of the size, location, and resolution of the display that are defined as parameters, we determine when a user is visually attending to elements of the display; and in particular, when they are looking at another person's avatar. When this occurs and fixation is prolonged, the social monitor increases the focus level on the observed party.

The OMNI Awareness View

One problem with existing systems is the difficulty of noticing when others gather information. Where social and peripheral attention processes mean that people constantly monitor their real world surroundings for the presence and gaze of others [41], limited screen space and window occlusion can limit attention to avatars or notifications online [38], and these displays can be distracting as well [12].

The OpenMessenger Notification and Interaction (OMNI) view is intended to exploit the properties of human visual perception in concert with the eye tracker social monitor and OMF. In the OMF, OMNI is classified as a *view*. The OMNI display consists of avatars representing contacts currently logged in, in addition to basic awareness information about these contacts. The display is projected onto a surface behind and above the user's primary monitor (or displayed on a large screen in the same space), so that it appears in the periphery of the user's visual field. Avatars are arranged in a semi-circle around the user's primary monitor, with their spatial arrangement initially determined by the order in which users login to the OM server, but this could be changed by the user to reflect natural groupings based on external factors such as project roles or relationship types.

OMNI has two primary uses: 1) displaying others' gathering behavior, and 2) facilitating gathering information about others.

Displaying. For displaying others' behavior, two visual parameters are manipulated: physical distance and motion. Physical distance of a contact's avatar from the user's body in the real world (presumed to be at the center of the display, though this could be altered) is inversely related to OM focus level. Contacts with a higher focus level on the current user (i.e., those that are gathering information) appear closer to the user. Distance between a user and a contact with focus level zero is defined by the vertical height of the display (i.e., the maximum possible distance given screen size constraints). Distance decreases by 20% of the overall distance with each increase in focus level until the avatars are touching at focus level five.

In line with findings from [6], we use motion to attract user attention in the event of change. As a contact increases their focus on the current user, the contact's avatar gradually moves toward the user on the OMNI display. Movement rate correlates directly with the rate of increase in focus level and occurs at a linear rate. As focus level decreases, the avatar moves away from the center at a rate slower than the approach. Movement away occurs more slowly to minimize distraction in the event that a contact looks back at an avatar soon after looking away.

Gathering. When the current user wishes to gather information about a contact, they focus on that contact's avatar (using a mouse, eye tracker, or other device with a social monitor). As they maintain focus on this avatar, more monitor information is displayed (see Figure 2) about that contact, reflecting the increase in focus level.

At this point, our social monitor and view architecture becomes quite powerful in supporting joint action and the notion of approach as a form of interaction. Consider the case where Alex wishes to interact with Bill. Alex looks at his OMNI display and sees that Bill is online. He then fixates on Bill's avatar to get more information (Figure 2) about him. As Alex gets more information (and his focus level on Bill increases), Alex's avatar will move toward Bill's on Bill's OMNI display. Bill notices this and looks at Alex's avatar, thus increasing Bill's focus level on Alex. As Alex's avatar moves toward Bill, more information about Alex (from his monitors) will

appear next to Alex's avatar on Bill's display. If Bill continues to look at this information, (i.e., as Bill's focus level on Alex increases), Bill's avatar on Alex's display moves toward Alex. Alex knows that Bill has noticed him, and can watch his display to see if Bill continues to attend to him. With sustained attention, they could opt to start a conversation or simply be aware of each other's interest.

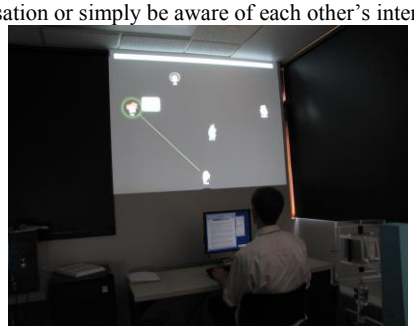

Figure 3. A user seated in front of an OMNI display, with four remote contacts. The user is focused on the leftmost contact, as indicated by the line.

Discussion

By focusing on both gathering and display of awareness information as interdependent components of the joint activity involved in negotiating mutual attention, OMF provides an extensible and general framework for experimenting with the management of interpersonal attention. It is the product of four years of development efforts, including three versions of OM. Here we discuss the implications of some key design decisions and assess the OMF's technical properties.

Sensor Data for Both Gathering and Display

We began by noting that prior systems often considered gathering and display in isolation, but that in reality these behaviors often correlate or even overlap. One unique and important attribute of the OMF is that data from sensors are used for both gathering *and* display. This stands in contrast to previous systems, which have largely treated sensor data primarily as a way to gather information about others (e.g., [7, 24]). OMF's sensor/monitor/focus level architecture is agnostic about who is "starting" an interaction or who is "observing" vs. "being observed." It uses sensor data to determine the focus of attention of *both* the observed party *and* the observer.

The implications of this approach are illustrated by the Eye Tracker sensor. If the eye tracker detects Alex looking at Bill's avatar, this single awareness event triggers both an increase in the amount of information Alex sees about Bill (gathering), as well as a notification to Bill that this is taking place (display). In this way, gathering and display are coupled more naturally, and it is technically possible to support the interactive aspects of awareness discussed by Schmidt [44].

More research and experimentation, however, are needed to implement mechanisms for interaction via gathering and display in ways that will not confuse or overwhelm users. This is the focus of our current and future research.

Approaching, Interacting and Privacy

Our second key point was that approach is a form of interaction, and people should be notified when others gather information about them. While the OMF supports notification for joint action purposes, this does not mean that notification will occur every time information is accessed. From a privacy standpoint, OMF's architecture supports but does not strictly enforce notification. Using Boyle's [15] distinction between confidentiality and solitude, we think about monitors as protecting the confidentiality of detailed sensor data, while focus levels protect solitude by reducing the amount of information that is displayed at any one time.

The combination of these mechanisms means that there are two potential cases where gathering could occur without notification. The first of these results from possible variation in views between users. It is possible that a user could implement or select a view that, for design or other reasons, does not display instances of certain types of gathering by other users. In these cases, the user would not be notified when these types of gathering took place. Second, receiver-side filtering of awareness events means that data are sent to clients regularly, but notification is provided to an observed user only when the data are accessed via a view. As data is stored on the client, it is theoretically possible that the user could hack the system for access in a way that does not result in notification. While these issues merit consideration, we do not feel they present a major threat to privacy because they seem unlikely to occur often in trusted groups of collaborators, and because the OMF provides other privacy controls that regulate what information is shared. Technical Assessment of OMF

Generality. The OMF is general in that it can theoretically handle any type of awareness-based attention management in which users gather information about the displayed activity or status of their collaborators, and adjust or update their own activities based on this information. The framework can easily be extended to incorporate data from any software or hardware sensor providing a data stream that can be parsed by an OMF monitor. This could include information to support nonverbal or semi-synchronous communication, or those that provide additional verbal/synchronous channels (e.g., audio, video).

The framework architecture also supports easy development of novel mechanisms for increasing or decreasing focus. While our sample implementation increases focus level when another user's avatar is the focus of attention, this could be extended to increase focus when users are focused on the same object (e.g., a document), or even when certain users are focused on each other.

Flexibility. The ideas behind OMF have been tested primarily with synchronous or semi-synchronous interaction, but the design of the framework is such that the architecture and general implementation approach could be adapted for use in other systems - other awareness / messaging systems, but also in a variety of other applications (e.g., shared editors, workflow systems, or document sharing).

The framework is flexible enough to be used in a variety of contexts, although we have primarily developed and tested it in a desktop/office environment. The general architecture, however, means that a group of users could be using individual clients with very different sensors, sensor managers, monitors and views; but these could all smoothly exchange data via the common structure of awareness events and focus level. In future we plan to adapt the system to mobile devices and sensors.

Performance. The nature of the OMF and current implementations means that the computation and bandwidth requirements for the framework are small. In particular, sensing, monitoring, and distribution of information functions of the OM framework are lightweight. We will consider performance further in future work.

Limitations and Future Work

There are several limitations that provide opportunities for substantial future work.
One clear limitation in our assessment is the lack of a field or laboratory evaluation, and this is part of our immediate future plans. We do not focus on this here, but instead point to several other key questions. First, we provide an operational and conceptual framework for supporting joint attention management, but leave aside substantial questions about how sensor data should be parsed, interpreted and aggregated as the number of sensors and granularity increase. Parsing sensor data has received substantial attention in recent years (e.g., [7, 25, 42]) and we aim to incorporate and build on these techniques. Relatedly, we currently use a naïve approach to increasing focus levels that assigns equal weight to different modes of input. We aim to develop a more sophisticated model, defined within OMF, that reflects the nuance of cues and modes of social attention.

Conclusion

We have presented a joint action approach to interpersonal attention management in supporting awareness and informal interaction. In our approach, actions are assumed to occur in anticipation of or in response to acts by others. Our OpenMessenger Framework provides operational solutions for the problems of: 1) discerning focus of attention, and treating this differently when focus is on another person; 2) allowing for joint action both during and prior to conversational interaction; and 3) allowing for awareness of other users' presence and behavior. The software and examples are available at: http://collabtech.hci.cornell.edu/projects/openmessenger.php so that other researchers can further test and explore the idea of joint attention.

Acknowledgements

We thank all who have helped develop OMF, especially Sigurd Teigen, Jeesung Na, Maryam Mustafa, Suchi Agicha, Oleg Krohkin, and Yang (Leon) Liu. This project is supported in part by the USDA Cooperative State Research, Education and Extension

Service (Hatch Project #NYC-131439), the US National Science Foundation (#IIS-0942659), and the Institute for the Social Sciences at Cornell University.

References

1. Ackerman, M. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human Computer Interaction*, 15, 2/3 (2000), 181-203.
2. Allen, T. J. *Managing the Flow of Technology*. MIT Press, Cambridge, MA, 1977.
3. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. Cambridge Press, Cambridge, UK, 1976.
4. Avrahami, D., Fussell, S. and Hudson, S. IM waiting: timing and responsiveness in semi-synchronous communication. In *Proc. ACM CSCW* (2008), 285-294.
5. Bailenson, J. N., Beall, A. C., Blascovich, J. and Turk, M. Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *Presence*, 13, 4 (2004), 428-441.
6. Bartram, L., Ware, C. and Calvert, T. Moticons: detection, distraction and task. *International Journal of Human-Computer Studies (IJHCS)*, 58(2003), 513-545.
7. Begole, J., Matsakis, N. and Tang, J. Lilsys: Inferring Unavailability Using Sensors. In *Proc. ACM CSCW* (2004), 511-514.
8. Benford, S. and Fahlen, L. A spatial model of interaction in large virtual environments. In *Proc. ECSCW* (1993), 109-124.
9. Birnholtz, J. Adopt, Adapt, Abandon: Understanding Why Some Young Adults Start, and then Stop, Using Instant Messaging. *Computers in Human Behavior*, 26, 6 (2010), 1427-1433.
10. Birnholtz, J., Gutwin, C. and Hawkey, K. Privacy in the open: how attention mediates awareness and privacy in open-plan offices In *Proc. GROUP 07* (2007), 51-60.
11. Birnholtz, J., Gutwin, C., Ramos, G. and Watson, M. OpenMessenger: Gradual Initiation of Interaction for Distributed Workgroups. In *Proc. ACM CHI* (2008), 1661-1664.
12. Birnholtz, J., Reynolds, L., Mustafa, M., Luxenberg, E. and Gutwin, C. Awareness Beyond the Desktop: Exploring Attention and Distraction with a Projected Peripheral-Vision Display. In *Proc. Graphics Interface* (2010),
13. Birnholtz, J. and Tang, D. Sharing Awareness Information Improves Interruption Timing and Social Attraction(in preparation).
14. Booker, J. E., Chewar, C. M. and McGrenere, J. Usability testing of notification interfaces: are we focused on the best metrics? *Proc. ACMSE*(2004), 128-133.
15. Boyle, M. and Greenberg, S. The Language of Privacy: Learning from Video Media Space Analysis and Design. *TOCHI*, 12, 2 (2005), 328-370.
16. Chau, M. and Betke, M. *Real Time Eye Tracking and Blink Detection with USB Cameras*. 2005, Boston University Computer Science Technical Report.
17. Clark, H. H. *Using language*. Cambridge University Press, New York, 1996.
18. Clement, A. Considering privacy in the development of multi-media communications. *Computer Supported Cooperative Work*, 2(1994), 67-88.
19. Dabbish, L. and Kraut, R. Controlling Interruptions: Awareness displays and social motivation for coordination. In *Proc. ACM CSCW* (2004), 182-191.
20. Davis, S. and Gutwin, C. Using Relationship to Control Disclosure in Awareness Servers. In *Proc. Graphics Interface '05* (2005), 75-84.
21. Dourish, P. and Belotti, V. Awareness and Coordination in Shared Workspaces. In *Proc. Proc. ACM CSCW* (1992), 107-114.
22. Dourish, P. and Bly, S. Portholes: Supporting awareness in a distributed work group. In *Proc. ACM CHI* (1992), 541-547.

23. Fish, R. S., Kraut, R. and Chalfonte, B. The VideoWindow system in informal communication. In *Proc. ACM CSCW* (1990), 1-11.
24. Fogarty, J., Au, C. and Hudson, S. Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition In *Proc. UIST* (2006), 91-100.
25. Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C. and Yang, J. Predicting Human Interruptibility with Sensors, *TOCHI* 12, 1 (2005), 119-146.
26. Fogarty, J., Lai, J. and Christensen, J. Presence versus Availability: The Design and Evaluation of a Context-Aware Communication Client. *International Journal of Human-Computer Studies (IJHCS)*, 61, 3 (2004), 299-317.
27. Forsyth, D. *Group dynamics*. Brooks/Cole, Pacific Grove, CA, 1998.
28. Frischen, A., Bayless, A. P. and Tipper, S. P. Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133, 4 (2007), 694.
29. Gaver, W., Moran, T., MacLean, A. and Lovstrand, L. Realizing a Video Environment: EuroPARC's Rave System. In *Proc. ACM CHI* (1992), 27-35.
30. Goffman, E. *The presentation of self in everyday life*. Anchor Books, New York, 1959.
31. Harrison, S. *Media space 20+ years of mediated life*. Springer, London, 2009.
32. Heath, C., Luff, P. and Sellen, A. Reconsidering the virtual workplace: flexible support for collaborative activity. In *Proc. ECSCW* (1995), 83-99.
33. Iqbal, S. T. and Bailey, B. P. Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. In *Proc. ACM CHI* (2007), 697-706.
34. Isaacs, E., Tang, J. and Morris, T. Piazza: A desktop environment supporting impromptu and planned interactions. In *Proc. ACM CSCW* (1996), 315-324.
35. Kastner, S. and Ungerleider, L. G. Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience*, 23(2000), 315-341.
36. Kendon, A. *Conducting interaction: patterns of behavior in focused encounters*. Cambridge University Press, Cambridge, UK, 1990.
37. Kraut, R., Egido, C. and Galegher, J. Patterns of Contact and Communication in Scientific Research Collaboration. In *Proc. ACM CSCW* (1988), 1-12.
38. McCrickard, D. S., Czerwinski, M. and Bartram, L. Introduction: design and evaluation of notification user interfaces. *Intl Journal of Human-Computer Studies*, 58(2003), 509-514.
39. McEwan, G. and Greenberg, S. Supporting social worlds with the community bar. In *Proc. ACM GROUP* (2005), 21-30.
40. McFarlane, D. C. and Latorella, K. A. The scope and importance of human interruption in human-computer interaction design. *Human Computer Interaction*, 17, 1 (2002), 1-61.
41. Nummenmaa, L. and Calder, A. J. Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13, 3 (2008), 135-143.
42. Olguin, D. O., Gloor, P. A. and Pentland, A. Capturing individual and group behavior with wearable sensors. *Proc. of the AAAI Spring Symposium on Human Behavior*(2009).
43. Rodden, T. Populating the application: a model of awareness for cooperative applications. In *Proc. ACM CSCW* (1996), 87-96.
44. Schmidt, K. The problem with 'awareness'. *Computer Supported Cooperative Work*, 11(2002), 285-286.
45. Sudnow, D. Temporal parameters of interpersonal observation. In D. Sudnow, ed. *Studies in Social Interaction.* Free Press, New York, 1972.
46. Tang, J. Approaching and leave-Taking: Negotiating Contact in Computer-Mediated Communication. *ACM TOCHI*, 14, 1 (2007), 1-26.
47. Whittaker, S., Frohlich, D. and Daly-Jones, O. Informal Workplace Communication: What is It Like and How Might We Support It? In *Proc. ACM CHI* (1994), 131-137.