In Proceedings of INTERACT'99. IFIP: Edinburgh, UK, 1999.

Designing Awareness with Attention-based Groupware

Roel Vertegaal
Cognitive Ergonomics Department
Twente University, The Netherlands
roel@acm.org

ABSTRACT A design rationale for the implementation of awareness features in the attention-based GAZE Groupware System is discussed. Attention-based groupware uses a framework for the design of awareness features based on the capturing, conveyance and rendering of information about human attention. The aim is to integrally provide information about the focus of conversational as well as workspace activities of participants. Our design themes were: implicit capturing of awareness information; scalability of networked awareness information; and representation of awareness information using natural affordances. Eye tracking provides a direct and noncommand way of capturing human attention. It allows attentive information to be conveyed separate from the communication signal itself, in a machine-readable format. This eases the integration of Conversational and Workspace Awareness information, and allows network bandwidth consumption of this information to scale linearly with the number of users. Attentional focus also provides an organizational metaphor for the rendering of awareness information. By combining a more strict WYSIWIS general communilaboration tool (a 3D virtual meeting room) with more relaxed-WYSIWIS focused collaboration tools (2D editors), the attention of human participants can be guided and represented from broad to focused activity.

KEYWORDS CSCW, groupware, videoconferencing, awareness, attention, gaze, eye tracking.

1. INTRODUCTION

Groupware systems have long suffered from the lack of an integral paradigm for the provision of awareness information. This has led to a plethora of user interface widgets for awareness support, each using its own metaphoric representation of awareness information (Greenberg, 1996, Gutwin et al., 1996a, Sohlenkamp and Chwelos, 1994). Vertegaal et al. (1997) presented a comprehensive framework for the design of awareness features based on the capturing, conveyance and rendering of information about human attention. In this approach, awareness is defined as *knowledge about the attention of others*. Their model focused on the provision of Micro-level Awareness information, which conveys the attention of users during synchronous distributed communication and collaboration (we will use the term *communilaboration* for the intersection of these two). Micro-level Awareness provides two kinds of information: information about whom participants are communicating with (Conversational Awareness), and about what they are working on (Workspace Awareness). Vertegaal (1999) showed how representations of visual attention in the form of gaze directional cues (relative body positioning; head orientation; and gaze at the facial region) are the most reliable non-verbal indicators of whom people communicate with (their *dialogic attention*) in multiparty face-to-face situations. More generally, we recognize the potential of human visual attention as a transparent and ubiquitous means for mediating awareness about other participants' attention for:

1) Persons;
2) Objects in a workspace;
3) The relation between these entities.

In this paper, we discuss our design rationale for the implementation of awareness features in the GAZE Groupware System (Vertegaal et al., 1998, Vertegaal, 1999), based on conveying the attention of others. First, we discuss the design constraints of the capturing, conveyance and rendering of human attention. Then, we will briefly discuss our prototype.

2. DESIGN RATIONALE

Our aim was to design a groupware system with integrated and transparent support for Micro-level awareness features. As a main functional requirement, our system was to provide a seamless integration between Conversational and Workspace Awareness (Buxton, 1992). The use of gaze directional cues in face-to-face conversations provided a paradigm on which such integration could be based. Our design strategy was motivated by the following themes:

1) *Implicit Collection of Awareness Information.* Rather than asking users to make explicit verbally or otherwise whom or what they are attending to, a clever monitoring of the spatial properties and timing of normal user behaviour (e.g., their system input) can provide a wealth of implicit information about their activities. We thus took a noncommand approach to providing awareness information, as discussed by Nielsen (1993). This should lead to a more transparent and efficient interface, with lower mental load and less interruption of task-oriented activities. In order to accomplish this in a mediated setting one does not necessarily need intelligent systems. All that is required is a paradigm for monitoring input activity of individual users, and presenting this as awareness information to users on the other side of a network.

2) *Scalability of Networked Awareness Information.* Since the purpose of groupware systems is to support many users, typically across a computer network, scalability of awareness information should be seen as an essential technical requirement. This is particularly true if one plans to use standard internet connections (Vertegaal and Guest, 1995).
3) *Representing Awareness Information With Natural Affordances.* According to Sohlenkamp and Chwelos (1994), the design of the *'Look and Feel'* (i.e., the perceived aspects of a user interface) of groupware applications should, where possible, be based on intuitions, knowledge and skills that people have acquired through years of shared work in the real world. Such knowledge may include the current Graphical User Interface (GUI) *Desktop Metaphor*, with its direct manipulation character. Gaze directional cues may provide us with a suitable metaphor for representing Conversational Awareness information. All that is required is an extension of this metaphor to the workspace, providing information about other users' relations to shared objects.

2.1 Implicit Collection of Awareness Information: Measuring Attention

We agree with Dourish and Bellotti (1992) that awareness information should be collected in a passive fashion, rather than being provided explicitly by participants. Nielsen (1993) describes a completely new user interface paradigm based on this principle: *noncommand interfaces*. According to Nielsen, noncommand interfaces, like face-to-face conversations, rely on a more fuzzy dialogue between users and user interfaces than is the case with current user interface paradigms. In the noncommand paradigm, instead of a user issuing commands (by means of a command line syntax or by clicking menus or icons with a mouse), the computer *observes* user activity. The system then tries to make sense of available human input using a set of heuristics or a disambiguation process which could be similar to *grounding* in human dialogue (Clark and Brennan, 1991). Thus, computers would only need to query the user when certain information, required to understand what action should be taken, is deemed missing. We believe noncommand interfaces, if applied appropriately, can lead to a more transparent and efficient interface, with lower mental requirements and less interruption of task-oriented activities. By means of anticipation and estimation, noncommand input may take us a step further towards the original goal of direct manipulation interfaces: the shifting of user attention from tool to task. In order to accomplish this in a mediated setting, we do not necessarily need intelligent systems. All that is needed is a specification of what individual user activity should be monitored, and how this should be presented as awareness information to users on the other side of a network. Based on our definition of awareness, what we should monitor is the locus and temporal pattern of individual users' attention (see (Vertegaal et al., 1997) for a discussion). Depending on the application, there are a number of ways in which such monitoring might be accomplished:

1) *Using Video Cameras.* A great benefit of video data for Micro-level Awareness purposes is its real-world and temporal nature. For example, video data may be very useful for conveying the attention span of others by means of their body movements, or real-world objects in the focus of other people's attention. A problem with video is that it can be difficult to achieve a seamless integration of spatial Conversational and Workspace Awareness properties (Buxton, 1992, Okada et al., 1994). If the shared workspace is displayed on computer screens, then depending on the positions of computer screens and the representation of work spaces on those screens, angles of looking or gesturing may easily become incoherent with actual participant attention. The problem of achieving eye-contact using camera/display units is good example of this issue (see (Vertegaal, 1998, Vertegaal, 1999) for a discussion). Another problem with video input is that the conversion of *generic* video images into a machine-readable format is still problematic. This problem may, for now, inhibit the use of such information by a noncommand interface for resolving decisions, e.g., about what awareness information to convey. A third problem with use of video data may be the heavy network bandwidth requirements, which we will discuss later.
2) *Using Microphones.* According to Vertegaal (1998), speech activity can be an excellent predictor of turntaking patterns. As such, data from individual users' microphones might be used to gauge Conversational Awareness information. However, microphone data may need disambiguation before being useful as a provider of awareness information, or as input data in a noncommand decision process. Too literal an interpretation of such information, for example, when determining whom people are listening to in multiparty communication, may therefore be detrimental to user performance (e.g., see (Buxton et al., 1997) for a discussion of problems with LiveWire voice-activated switching). Again, the temporal properties of audio data seem the most relevant. Microphone input could, for example, be used to monitor user presence or activity. Microphone input seems less appropriate for providing Workspace Awareness information. As for network constraints, audio data requires far less bandwidth than video data. In addition, we believe the availability of speech should be regarded a minimum requirement during synchronous mediated communilaboration anyway (Chapanis, 1975).
3) *Using Manual Input Devices.* Manual input devices such as the mouse and keyboard are important means for gauging Micro-level Awareness information. In text-based environments, the duration and aim of keyboard input may provide Conversational Awareness information in ways similar to the above use of microphone input. Attention of users could then be represented by, e.g., font size of textual communication. In graphical user interfaces, a representation of the location of pointing devices within a shared workspace may be used to convey Workspace Awareness information. Many current-day groupware systems already provide such *telepointers* as an indication of the locus of participant activity (Gutwin et al., 1996b).

Advantages of the use of manual input devices for providing awareness information include: they are low-cost and ubiquitous; they already are the main means of manipulating objects in shared workspaces; their data is machine-readable and low-bandwidth by nature. A disadvantage of manual pointing devices may be that they often do not return to a zero state. If a participant leaves her mouse pointer at a position within a shared workspace, the telepointer representation may falsely indicate her attention to that part of the workspace. In the future, such problems might be circumvented by basing the decision to represent a telepointer on a fuzzy assessment of data from different input devices. We believe a more important restriction in the use of manual pointing devices is that they typically require an explicit manipulative action. Hence, they seem suitable mostly for gauging action-related awareness information, such as conveying the direct manipulation of shared objects. The use of manual pointing in providing Conversational Awareness information seems limited to manual deixis towards other participants.

4) *Using the Real World as an Input Device.* A recent development is the use of real-world objects, rather than software objects, as a user interface to software processes (so-called Tangible Media, see (Ishii and Ullmer, 1997)). In this approach, the orientation and position of objects in the real world, e.g., on a desk, is gauged by means of sensors or simple image recognition techniques (for example, by recognizing barcode stickers on objects (Underkoffler and Ishii, 1998)). Attributes of real objects could thus provide low-bandwidth Workspace Awareness information to participants on the other side of a network, where they could be re-synthesized by projection onto their desk. The biggest advantage of this approach is the richness and transparency of the interface for single users. For now, the biggest drawback is that software manipulation of real-world artifacts is still limited. Thus, the joint manipulation of real objects may be problematic. Although we recognize the potential of this technique, we consider it beyond the scope of this paper. In a related approach, data-suits and other forms of sensor technology may gauge a wide range of parameters of human behaviour in various forms of transparency, such as head or body orientation (Rabb et al., 1979). Eye and head orientation tracking are examples of such technology. These techniques would seem the most relevant in this category for comprehensive gauging of awareness information in general, and Conversational Awareness information in particular.

5) *Using Eye and Head Tracking Devices.* The orientation of the human eye or head can be gauged by tracking devices. Although at the moment, eye tracking technology is not yet used for generic input purposes, this is changing rapidly (Joch, 1996, Nielsen, 1993). Capturing the actual focus and span of visual attention by means of an eyetracking system may provide a relatively direct and high-resolution means of capturing information about participants' attention for actions, objects and people alike. Eye input may thus provide an integrated approach for gauging Conversational and Workspace Awareness information. In addition, eyetracker information is machine-readable, low-bandwidth and noncommand by nature (Nielsen, 1993). Many problems with the application of eyetracking in user interfaces were in fact *due* to inadvertent use of eye fixation information for issuing system commands (the *"Midas Touch"* problem, see (Velichkovsky et al., 1997)). A clear disadvantage of eye input is that eyetracking devices are still rather expensive. However, this seems mostly due to the low production volume. Indeed, low-resolution eyetrackers are already becoming available for less than $1500. Unfortunately, eyetracking still has an undeserved negative reputation in terms of usability. Archaic requirements such as bulky head attachments or fixation of the user's head need no longer apply. With up to 900 cm^3 of head movement tolerance, the transparent application of desk mounted eyetrackers for desktop computer input purposes has recently become a realistic option (Applied Science Laboratories, 1998, LC Technologies, 1997). It is with ranges larger than these that head orientation sensors become a good alternative, at least for gauging Conversational Awareness information (Rabb et al., 1979). The inaccuracy of head orientation information would probably require an alternative source of input for the measurement of Workspace Awareness information.

Next, we will discuss the impact of the selection of input modality on network bandwidth requirements of a groupware system.

2.2 Scalability of Networked Awareness Information

Since the purpose of groupware systems is to support many users, in our case across a computer network, scalability of the network bandwidth consumed by awareness functionality is a technical design constraint that should be taken seriously (Greenhalgh and Benford, 1995, Vertegaal and Guest, 1995). We will limit our discussion to a simple comparison between the impact on network resources of methods of input for Conversational Awareness information. From the above discussion, it becomes apparent that currently, the most integral candidates for gauging Conversational Awareness information are *1) video cameras* and *5) eye or head tracking devices*.

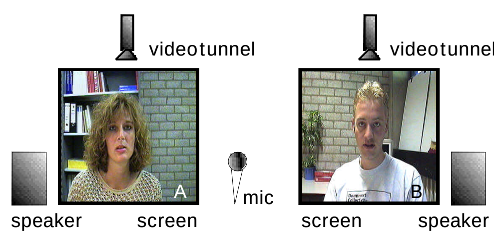

Figure 1. *Video tunnel setup. Each participant is represented by a camera/display unit.*

System 1: Using Video Cameras

As discussed, Conversational Awareness can effectively be constituted by mediation of information about the relative position, head orientation and gaze of individual users

(Vertegaal, 1999). When video cameras are used to capture this information, it is advisory to use a multiple camera setup, such as the one depicted in Figure 1. Each participant has a camera/display setup for each other participant using the system. In between the camera and display of each unit, a half-silvered mirror is placed at an angle of 45 degrees. This *video tunnel* principle allows gaze at the facial region, at least to some extent, to be conveyed (Acker and Levitt, 1987). A good example of a mediated system using such setups is *MAJIC* (Okada et al., 1994). Although there are problems with the consistent preservation of eyecontact using such systems (Vertegaal, 1998, Vertegaal, 1999), there are also problems with the network load generated by such systems. In normal packet-switched networks (such as the Internet), the video from each camera and audio data from the microphone in the above system would need to be broadcast individually to each other participant in a meeting (rather like in a Cable TV network). *Multicasting* is a new Internet technique which prevents the inefficient use of the network bandwidth caused by such individual broadcasting techniques (Ericksson, 1994, Vertegaal and Guest, 1995). In Multicasting, each unique stream of video data is put on the net only once, and is then picked up by the system of each other participant in the meeting (rather like a standard TV broadcast is picked up by the TV antenna of viewers that are tuned in). Thus, using Multicasting, the total bandwidth consumption for System 1 would be equal to:

$$B = nA + n(n-1)V \qquad \text{(Equation 1)}$$

In this equation, B is the total amount of bandwidth used, n is the number of participants, A is the amount of bandwidth per audio input, and V the amount of bandwidth used per unique video stream. It is clear that System 1 does not scale linearly with the number of participants. With four participants, 12 units of video bandwidth are required. With six participants, this rises to 30 units of video bandwidth.

System 2: Using Eye or Head Tracking Devices

When eye or head tracking devices are used, the manipulation of images of individual users could be used to convey Conversational Awareness information to other participants (Vertegaal, 1999). In such system, pictorial representations of users would be manipulated such that their relative positioning, head orientation and gaze would be preserved (Vertegaal, 1997). This manipulation could occur according the measured locus of visual attention. An example of a system in which head trackers are used to convey head orientation is the *Talking Heads* system by (Negroponte). When motion video would be conveyed using such systems, the total bandwidth consumption B in a Multicast network equals:

$$B \approx nA + nV \qquad \text{(Equation 2)}$$

Since video is not used to convey visual attention *itself*, this system scales linearly with the number of participants. When still pictures are used, V approaches zero, since all that is conveyed is the coordinates of visual attention of the participants. In that case, the amount of bandwidth needed is no more than would be required by audio only.

Concluding Remarks

Use of video input for conveying Conversational Awareness information simply does not scale well with the number of participants. This, and the fact that the availability of a measure of visual attention should ease the integration of Conversational and Workspace Awareness information, led us to prefer System 2 with eyetracker input as a basis for our system design. Given the spatial range of available eyetracking devices, we, for now, limited our design to a desktop computer environment. We decided to initially build an audio-mediated environment in which still images of participants are manipulated in order to visually represent Conversational Awareness information. Empirical evidence shows that still images can successfully convey such information (see (Vertegaal, 1999) for a discussion). Next, we will discuss design issues in the representation of awareness information.

2.3 Representing Awareness: Designing a Virtual Meeting Room

In our discussion of the design of representations for groupware system functionality, we will concentrate on how Micro-level Awareness information could be represented in an audio-visual desktop computer environment. We will focus on the integral and synchronous provision of Conversational and Workspace Awareness information, rather than on the design of the communication and collaboration tools themselves.

This design theme relates to the “Look and Feel” of awareness functionality, guiding how the perceived aspects of a groupware system awareness interface could be rendered. In doing so, we wanted to make use of existing knowledge and skills of users as much as possible. We therefore chose a metaphoric design approach, in which elements of the interface and their behaviour would be based on real world equivalents as much as possible. We tried to use Gibsonian affordances to render awareness functionality into the user’s perception as directly as possible. Thus, we tried to allow users to rely as much possible on knowledge in the system image, rather than on knowledge in their heads. We agree with Sohlenkamp and Chwelos (1994) that a metaphoric design approach should not be followed too rigidly in order to prevent the inadvertent modeling of limitations inherent to the real world. Instead of modeling the real world on a one-to-one basis, we therefore attempted to model the essential bits only. Finding a basis for our representations in the real world included making use of users’ knowledge of current Graphical User Interface *Desktop Metaphor* (Smith et al., 1982). In the design of their DIVA groupware system functionality, Sohlenkamp and Chwelos (1994) simply expanded the single-user desktop paradigm to include multiple users, adding the elements *people* and *rooms* to elements already present in the desktop paradigm: *documents*, *desks* and *pointers* (with the latter becoming telepointers). Thus, they padded an existing computer metaphor with elements borrowed from the real world. In order to achieve a seamless integration of representations for Conversational and Workspace Awareness information (our main functional requirement), different user interface elements should have some form of spatial and temporal relation with each other. We therefore followed an approach similar to Sohlenkamp and Chwelos, building a *virtual meeting room* in which the above user interface elements are jointly represented (Ensor, 1998). However, as we will now discuss, a virtual meeting room alone may not be sufficient for supporting focused collaboration.

General versus Focused Collaboration

Stefik et al. (1987b) proposed the *"What You See Is What I See"* (WYSIWIS) paradigm as a means of providing a consistent and coordinated display of user interface elements to all participants. In strict WYSIWIS, all participants essentially have exactly the same display containing exactly the same information at exactly the same moment in time. In their Colab environment Stefik et al. (1987a, 1987b) supported WYSIWIS by maintaining synchronized views, and by offering facilities for telepointing with publicly visible cursors. This would allow participants to have a common understanding of their virtual world, permitting them to rely on the availability of external context in, for example, deixis. However, Stefik et al. (1987a) also pointed out that a strict application of WYSIWIS throughout user interface elements may be too inflexible. It may, for example, lead to problems in supporting the parallel work on different tasks by subgroups, or the transfer of information between private and public spaces. Instead, they recommended strict WYSIWIS as a foundational abstraction, with a selective easing of compliance (relaxed-WYSIWIS) along four dimensions:

1) *Display space.* Strict WYSIWIS applies to everything on an individual display; applying it only to a subset of visible objects (e.g., windows and cursors) relaxes this constraint.
2) *Time of display.* Strict WYSIWIS requires that images be synchronized; allowing delays in updating or viewing of images relaxes this constraint.
3) *Subgroup population.* Strict WYSIWIS requires shared viewing to apply to everyone in the full meeting groups; allowing sharing to be limited to subgroups relaxes this constraint.
4) *Congruence of view.* Strict WYSIWIS requires that images be identical; allowing alternative views relaxes this constraint.

As a response to this, Gutwin et al. (1996b) warned that relaxed-WYSIWIS may actually lead to a lack of awareness, since increased individual control reduces the group focus inherent in strict WYSIWIS systems. Thus, there may be a conflict between requirements for general and focused collaboration. We therefore decided to have a more strict environment for general group activities, and a more relaxed environment for focused collaboration activity. The virtual meeting room would provide a place on the display where general group activity is grounded in a rather strict manner. Within it, only congruence of view would be relaxed, and only in that each participant's viewpoint would be strictly located at the position of his representation. This way, we ensured effective application of gaze as a metaphor for conveying Conversational Awareness information. Individual viewpoints would otherwise be fixed such that all awareness information would be within field of view of all participants. On the rest of the display, around the virtual meeting room, focused collaboration could take place using task-specific relaxed-WYSIWIS document editors (e.g., those proposed by (Gutwin et al., 1996a, Greenberg, 1996)). Our WYSIWIS relaxation requirements for focused document editing were based on recommendations made by Baecker et al. (1993), and almost the opposite of those of the virtual meeting room. Different documents may appear at different locations on displays of individual participants; updating of images may depend on where individuals work within the document; document contents need be displayed only to the subgroup working on them; document contents is typically viewed from the same angle by all participants (e.g., during text editing), but position of individual users within documents (i.e., the part of the document that is displayed) is totally relaxed. However, in order to provide a common glue between document editors (focused collaboration tools) and the virtual meeting room (the general communilaboration tool) we introduced one constraint: there should always be at least one WYSIWIS representation of a telepointer linking the attention of a participant in the meeting room to his attention to sections of document content.

Attentional Focus as an Organizational Metaphor

In a larger perspective, the concept of *attentional focus* can be regarded as an organizational metaphor throughout the design, gluing representations of awareness functionality at different levels of refinement together so that they can be recognized as a whole. This becomes apparent when we consider the suggested user interface elements as a way of representing attention of participants. *Rooms* are ways of organizing the presence of people, signalling the general availability of their attention for a common communilaborative goal. Within rooms, the co-location and orientation of *persons* is a way of organizing the joint attention of sub-groups towards a common communilaborative task (Nakanishi et al., 1996). *Desks* are ways of organizing task-specific objects, providing an overview of their availability for collaborative attention. Within desks, *documents* signal the availability of a task as a focus for collaborative attention. Also within desks, *telepointers* signal the actual focus of collaborative attention towards a certain task. Finally, when documents are opened, relaxed WYSIWIS document editors allow participants to focus their attention according to individual interest. As will be discussed, telepointers link this focus within the document to the focus within in the virtual meeting room (as an example, see the Gestalt view of the SASSE environment (Baecker et al., 1993)). Thus, using the above organizational metaphors, the focus of attention of participants may be described and guided from the very general to the very specific.

Visual Representation of Interface Elements

We will now discuss how the above discussed user interface elements could be rendered with visual behaviour. To keep the user interface as simple as possible, we chose to represent only five elements of real meeting rooms:

rooms, *desks*, *persons*, *documents*, and *pointer light spots*. Other attributes include a stationary pad, exit sign, and a trash can.

1) *Rooms.* Rooms contain all people, desks and documents required for a synchronous distributed communilaboration session. Depending on the environment, rooms could be represented by text windows (e.g., in chat environments), 2D surfaces (e.g., DIVA (Sohlenkamp and Chwelos, 1994)), or 3D worlds (e.g., MASSIVE (Greenhalgh and Benford, 1995)). As will be discussed later, we wanted to use head orientation as a metaphor for conveying visual attention of participants. If this information was to be used for conveying Conversational Awareness as well as Workspace Awareness, 3D orientation would seem a requirement. We therefore chose a 3D room design, which could function as a container for organizing the attention of participants at the *presence* level. Just like people located in the same real room are able to see and hear each other, so too would people within a our virtual room hear and see each other. As with DIVA (Sohlenkamp and Chwelos, 1994), the entering of a person into a room could establish audio-visual communication links with people already present. We restricted ourselves to using rooms as a means for organizing private meetings only (i.e., a virtual meeting room (Ensor, 1998)).
2) *Desks.* These containers represent a way of organizing attention of participants towards any number of collaborative objects. Depending on the environment, a simple representation of a directory structure might be used for grouping shared files. However, we chose the single-user desktop metaphor as a basis, expanding it into a shared surface onto which iconic representations of shared file objects could be placed and organized by position (Smith et al., 1982). However, our representation would function not just as a means for organizing collaborative objects, but also as a means of organizing persons. By placing representations of persons around a 2D desk surface in our 3D meeting room, face-to-face round-table communilaboration could be used as a metaphor for integrating Conversational and Workspace Awareness information.
3) *Persons.* A participant is represented by a *persona*: a metaphoric rendering of real participant behaviour (Negroponte). An important functional requirement for personas is that they represent a participant's visual attention (Vertegaal, 1999). Although, depending on the environment, personas may be rendered by a name (chat environments), a 3D model (avatar environments), or a video stream (video conferencing), this rendering would need to include a visual representation of real participant attention towards other persons. Gaze may be considered an ideal metaphor for this purpose. Real images of participant gaze are the most effective way of conveying this (Vertegaal, 1999, Vertegaal, 1998). As discussed, we chose to initially base our design on still images, rather than motion video images. This to circumvent problems of parallax between camera position and participant representation, which, even when video tunnels are used, may impair correct perception of visual-attentive cues (Vertegaal, 1999, Vertegaal, 1998). In order to achieve a smooth integration of the persona in the 3D meeting room, we decided against the use of different images for conveying different loci of visual attention (Tanaka et al., 1996). Instead, for each participant, we suspended a single frontal snapshot — made while looking into the camera lens — in the 3D meeting room. 3D orientation of this 2D persona would then metaphorically convey the direction of gaze of that participant, as measured by the eyetracking device.
4) *Documents.* These containers represent a way of organizing attention of participants towards a particular task. In standard desktop environments, document icons typically function as a representation of associated document content (Smith et al., 1982). We took a similar approach. Document icons can be placed on a desk in the virtual meeting room as a means of sharing the associated content. This content can be accessed by opening the document icon (e.g., by double-clicking it), at which moment it is downloaded and displayed in a focused collaboration editor outside the virtual meeting room. Document editors appear only to those participants that opened the document, but the associated document icon on the desk remains visible to all. Documents can be associated with local editors, or editor software could be embedded as part of the document content. In principle, documents can contain any kind of information, as long as an associated editor is available to all parties. As discussed, information display in such editors would typically be based on all participants having the same point of view, but should otherwise follow a relaxed-WYSIWIS paradigm. As discussed, telepointers provide ways of linking the focus of attention of individuals on sections of document content to the document representation in the virtual meeting room.
5) *Telepointer Light Spots.* These represent the actual attention of participants for objects in a shared work space. During presentations by an individual in a group meeting, light spots produced by laser pointing devices are now widely used to communicate the exact focus of attention of a presenter. We used these light spots as a metaphor for telepointing, illuminating objects in a shared workspace according to the attention of individual participants. As a source of information about this attention, we could use mouse position or the actual point of gaze as provided by the eyetracker. With the latter, participants need not take any action other than looking to provide others with Workspace Awareness information. During general communilaboration in the meeting room, when a participant looks at a location on a desk, a light — appearing to be emitted from his persona — illuminates the spot. We thus borrowed a functional metaphor from the helmets used by miners to illuminate their work environment (the *Miner's Helmet* metaphor). The light spot is also associated with the emitting persona by means of colour coding. Multiple light spots of the same colour can be used to represent the same focus of attention at different levels of refinement. During focused collaboration in a document editor, when a collaborator looks at a location within the document content, a light with her colour illuminates the

spot. This light spot is visible only to persons working within the document. Therefore, whenever a person looks at document content, the associated document icon in the meeting room should also be illuminated by a light spot of his colour. This light spot is visible to all. If documents contain multiple sections, multiple light spots of the same colour could indicate in a strict-WYSIWIS fashion which section each collaborator is focusing on (see the Gestalt viewer of the SASSE environment (Baecker et al., 1993) as an example of how this might be accomplished). All light spots generated by a single persona were to remain tightly associated by movement and colour. This way, light spots may provide a kind of attentional glue between focused collaboration and general communilaboration activities.

3. THE GAZE GROUPWARE SYSTEM

Based on the above rationale, we developed a prototype groupware system which provides integral support for Conversational and Workspace Awareness by conveying the participants' visual attention. Instead of using multiple streams of video for this purpose, the GAZE Groupware System (GGS) measures directly where each participant looks by means of an advanced desk-mounted eyetracking system. The system represents this information metaphorically in a 3D virtual meeting room and within shared documents. The system does this using the Sony Community Place (Sony, 1997) plug-in, which allows interactive 3D scenes to be shared on a web page using a standard multiplatform browser such as Netscape. In this prototype, we did not yet integrate support for multiparty audio communication. Instead, the GAZE Groupware System can be used in conjunction with any multiparty speech communication facility such as an Internet-based audio conferencing tool, or standard telephony.

3.1 A Session in the GAZE Virtual Meeting Room

The GAZE Groupware System simulates a four-way round-table meeting by placing a 2D image (or persona) of each participant around a desk in a virtual room, at a position that would otherwise be held by that remote participant. Using this technique, each person is presented with a unique view of each remote participant, and that view emanates from a distinct location in space. Each persona rotates around its own *x* and *y* axes in 3D space, according to where the corresponding participant looks. Figure 2 shows the system in use in a four-way situation. When Robert looks at Roel, Roel sees Robert's persona turn to face him. When Robert looks at Harro, Roel sees Robert's persona turn towards Harro. This should effectively convey whom each participant is listening or speaking to. When a participant looks at the shared desk, a light spot is projected onto the surface of the desk, in line with her persona's orientation. The colour of this light spot is identical to the colour of her persona. This allows a participant to see exactly where the others are looking within the shared workspace. By direct manipulation, e.g., with their mouse, participants can put document icons, representing shared files, on the desk. Whenever a participant looks at a document icon or within the associated file, her light spot is projected onto that document icon. This allows people to use deictic references for referring to documents (e.g., *"Here, look at these notes"*). Shared documents are opened by double clicking their icon on the desk. When a document is opened, the associated file contents appears in a separate frame of the web page (see Figure 2). In this frame, an editor associated with the file runs as an applet. When a participant looks within a file, all participants looking inside that file can see a light spot with her colour projected over the contents. This light spot shows exactly what this person is reading. Again, this allows people to use deictic references for referring to objects within files (e.g., *"I cannot figure this out"*). For a more complete and technical description of the GAZE system, see (Vertegaal, 1999).

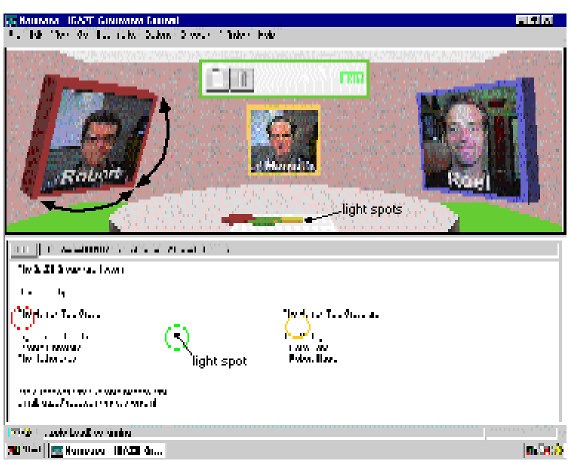

Figure 2. *The GAZE virtual meeting room (top) with a shared document editor (below).*

4. CONCLUSIONS

In this paper, we discussed a design rationale for the implementation of awareness features in the attention-based GAZE Groupware System. Attention-based groupware uses a framework for the design of awareness features based on the capturing, conveyance and rendering of information about human attention. Our main functional requirement in the design process was to achieve a seamless integration between Conversational Awareness information (whom are participants communicating with?), and Workspace Awareness information (what they are working on?). Our design themes were: the implicit collection of awareness information; the scalability of networked awareness information; and the representation of awareness information using natural affordances. Eye tracking devices provide a direct and noncommand way of capturing multi-modal human attention. Their main advantage is that they allow attentive information to be conveyed separate from the communication signal itself, in a machine-readable format. Not only does this ease the integration of Conversational and Workspace Awareness information, it also allows networked bandwidth consumption of this information to scale linearly with the number of users. We have shown how attentional focus may also provide an elegant organizational metaphor for the rendering of awareness information. By combining a more strict WYSIWIS general communilaboration tool (a 3D virtual meeting room) with more relaxed-WYSIWIS focused collaboration tools (2D WIMP editors), the attention of human participants can be guided and represented from broad to focused activity. Thus, the attention-based groupware paradigm provides an integral approach to the capturing and guidance of human awareness of others.

5. REFERENCES

Acker, S. and Levitt, S. (1987) *Designing videoconference facilities for improved eye contact. Journal of Broadcasting & Electronic Media,* **31,** 181-191.

Applied Science Laboratories (1998) *ASL Eyetracking Systems, http://www.a-s-l.com*, Bedford, MA USA.

Baecker, R. M., Nastos, D., Posner, I. R. and Mawby, K. L. (1993) *The User-Centred Iterative Design of Collaborative Writing Software.* In *Proceedings of INTERCHI '93*. IOS Press, Amsterdam. pp. 399-405.

Buxton, W. A. S. (1992) *Telepresence: Integrating Shared Task and Person Spaces.* In *Proceedings of Graphics Interface '92,* pp. 123-129.

Buxton, W. S., Sellen, A. J. and Sheasby, M. C. (1997) *Interfaces for Multiparty Videoconferences.* In *Video-Mediated Communication* (Eds, Finn, K. E., Sellen, A. J. and Wilbur, S. B.) Laurence Erlbaum Associates, Mahwah, NJ USA, pp. 385-400.

Chapanis, A. (1975) *Interactive Human Communication. Scientific American,* **232,** 36-42.

Clark, H. H. and Brennan, S. E. (1991) *Grounding in Communication.* In *Socially Shared Cognition* (Eds, Resnick, L. B., Levine, J. and Behreno, S. D.) American Psychology Association, Washington.

Dourish, P. and Bellotti, V. (1992) *Awareness and coordination in shared workspaces.* In *Proceedings of CSCW '92.* ACM, NY, Toronto, Canada, pp. 25-38.

Ensor, J. R. (1998) *Virtual Meeting Rooms.* In *Video-Mediated Communication* (Eds, Finn, K. E., Sellen, A. J. and Wilbur, S. B.) Laurence Erlbaum Associates, Mahwah, NJ USA, pp. 415-434.

Ericksson, H. (1994) *MBONE: The Multicast Backbone. Communications of ACM,* **37,** 54-60.

Greenberg, S. (1996) *A Fisheye Text Editor for Relaxed WYSIWIS Groupware.* In *Conference Companion of CHI '96.* ACM, Vancouver, Canada, pp. 212-213.

Greenhalgh, C. and Benford, S. (1995) *MASSIVE: A Collaborative Virtual Environment for Teleconferencing. ACM Transactions on Computer-Human Interaction,* **2,** 239-261.

Gutwin, C., Greenberg, S. and Roseman, M. (1996a) *Workspace Awareness Support With Radar Views.* In *Conference Companion of CHI '96.* ACM, Vancouver, Canada, pp. 210-211.

Gutwin, C., Roseman, M. and Greenberg, S. (1996b) *A Usability Study of Awareness Widgets in a Shared Workspace Groupware System.* In *Proceedings of CSCW '96.* ACM SIGCHI, Boston, MA USA, pp. 258-268.

Ishii, H. and Ullmer, B. (1997) *Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms.* In *Proceedings of CHI '97.* ACM, Atlanta, GA, pp. 234-241.

Joch, A. (1996) *What Pupils Teach Computers. BYTE,* **21,** 99-100.

LC Technologies, I. (1997) *The Eyegaze Communication System, http://www.lctinc.com*, Fairfax, VA.

Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T. (1996) *Freewalk: Supporting Casual Meetings in a Network.* In *Proceedings of CSCW '96.* ACM, Cambridge, MA., pp. 308-314.

Negroponte, N. *Talking heads-display techniques for persona. Unpublished paper. MIT Architecture Machine Group.* Cambridge, MA.

Nielsen, J. (1993) *Noncommand User Interfaces. Communications of ACM,* **36,** 83-99.

Okada, K.-i., Maeda, F., Ichikawaa, Y. and Matsushita, Y. (1994) *Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design.* In *Proceedings of CSCW '94.* ACM, Chapel Hill, NC., pp. 385-393.

Rabb, F., Blood, E., Steiner, R. and Jones, H. (1979) *Magnetic Position and Orientation Tracking System. IEEE Transactions on Aerospace and Electronic Systems,* **15,** 709-718.

Smith, D., Irby, C., Kimball, R., Verplank, B. and Harslem, E. (1982) *Designing the Star user interface. BYTE,* **7**.

Sohlenkamp, M. and Chwelos, G. (1994) *Integrating Communication, Cooperation and Awareness: The DIVA Virtual Office Environment.* In *Proceedings of CSCW '94.* ACM, Chapel Hill, NC, pp. 331-343.

Sony (1997) *Sony Community Place, http://vs.spiw.com/vs/.*

Stefik, M., Bobrow, D. G., Foster, G., Lanning, S. and Tatar, D. (1987a) *WYSIWIS Revised: Early Experiences with Multiuser Interfaces. ACM Transactions on Office Information Systems,* **5,** 147-167.

Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S. and Suchman, L. (1987b) *Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings. Communications of ACM,* **30,** 33-47.

Tanaka, S., Okada, K.-i., Kurihara, S. and Matsushita, Y. (1996) *Desktopconferencing System Using Multiple Still-Pictures:Desktop-MAJIC.* In *Proceedings of CSCW '96.* ACM, Cambridge, MA.

Underkoffler, J. and Ishii, H. (1998) *Illuminating Light: An Optical Design Tool with a Luminous-Tangible Interface.* In *Proceedings of CHI '98.* ACM, LA, CA, pp. 542-550.

Velichkovsky, B., Sprenger, A. and Unema, P. (1997) *Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem".* In *Proceedings of INTERACT '97.* IFIP, Sydney, Australia.

Vertegaal, R. (1997) *Conversational Awareness in Multiparty VMC.* In *Extended Abstracts of CHI '97.* ACM, Atlanta, GA, pp. 6-7.

Vertegaal, R. (1998) *Look Who's Talking to Whom.* PhD Thesis. Cognitive Ergonomics Department, Twente University, Enschede, The Netherlands.

Vertegaal, R. (1999) *The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration.* In *Proceedings of CHI'99.* ACM, Pittsburgh, PA.

Vertegaal, R. and Guest, S. (1995) *Network Issues in the Growth and Adoption of Networked CSCW Services. SIGCHI Bulletin,* **27,** 63-68.

Vertegaal, R., Velichkovsky, B. and Veer, G. V. d. (1997) *Catching the Eye: Management of Joint Attention in Cooperative Work. SIGCHI Bulletin,* **29**.

Vertegaal, R., Vons, H. and Slagter, R. (1998) *Look Who's Talking: The GAZE Groupware System.* In *Summary of CHI '98.* ACM, Los Angeles, CA.