Catching the Eye: Management of Joint Attention in Cooperative Work

(Appeared in SIGCHI Bulletin (29)4. ACM SIGCHI, 1997)

Roel Vertegaal, Boris Velichkovsky and Gerrit van der Veer

Abstract

In this paper, we show how different elements of awareness information in groupware systems can be defined in terms of conveying attentive states of the participants. Different kinds of awareness are distinguished: at macro- and micro-level, the latter consisting of workspace awareness and conversational awareness. We summarize the functional elements of micro-level awareness, organizing them hierarchically in terms of their relation to the attention of participants. We further discuss how groupware systems can capture and represent awareness by means of attention-based metaphors, and give an example of a virtual meeting room in which the gaze direction of the participants is conveyed by means of modern 'imaging' eyetracking technology.

Introduction

Groupware and videoconferencing systems allow groups of people to collaborate and communicate synchronously and interactively while at different locations. Current systems allow participants to interact by means of audio and video, allowing them to hear and see each other. However, a rather low level of acceptance of such systems [LOR] implies that important aspects of human communication are not supported in contrast to their presence in face-to-face interaction: in particular, we feel that people need to be better *aware* who is talking to whom and about what [VER].

Structuring Awareness

We feel it is time to start organizing different aspects of awareness into an analytical framework. This in an attempt to put a hold on the proliferation of terms indicating similar concepts (ironically, we will suggest new jargon in an attempt to get rid of the old), and the habit of defining awareness in terms of the GUI widgets which constitute it. Hopefully, such a framework will make it easier for human factors designers to structurally develop awareness *functionality* within their application framework. We will attempt to put some functional elements which we consider important into a framework based on the definition of awareness (within the realm of synchronous interactive systems) in terms of conveying the attention of others. A two level split is considered: macro-level awareness dealing with aspects of the world outside a virtual meeting, and micro-level awareness dealing with awareness aspects of a virtual meeting. We will then concentrate on micro-level awareness which in its turn is divided into workspace and conversational awareness.

Awareness: Towards Conveying Joint Attention States

We propose to define elements of awareness in terms of the time and place of the attention of other participants. Thus, we can look at awareness in communication and collaboration (we pitch the term *communilaboration* for the intersection of these two) in terms of a network of joint attention states. Once awareness is modeled in terms of attentive states of the participants, we can attempt to systematically capture and explicitly represent these attentive states in order to provide comprehensive awareness information. Before discussing a possible mapping of attentive states with awareness

information, we would first like to narrow our focus by defining complementary levels of awareness information.

Macro-level Awareness

Macro-level awareness are forms of awareness which convey background information about the activities of others prior to or outside of a meeting. This relates to *informal awareness* [GRE] and *general awareness* [GAV]. Both are defined as "…the general sense of who is around and what others are up to". Who is available for a meeting, what will the meeting be about, where, why and when will it take place and what tools will be used? Most of this information is rather discrete by nature. Often, small, low-frequency images [DOU] or activity indicators [GRE] can be used to sense the availability of persons for communilaboration. In this paper, however, we would like to concentrate on a relatively neglected issue of micro-level awareness and how it can be constituted by representing the attention of others.

Micro-level Awareness: Conversation and Workspace

Micro-level awareness are forms of awareness which give online information about the activities of others during the meeting itself. This relates to the concept of Focused Collaboration Awareness discussed by Gaver [GAV]. Micro-level awareness usually has a more continuous nature than its macro-level counterpart. It consists of two categories: *Conversational awareness* and *Workspace awareness*. Conversational awareness contains information about who is communicating with whom, workspace awareness contains information about who is working on what. Both imply a notion of space: in order to constitute these forms of awareness, one needs to know where 'who' is and where 'what' is. Together, they can provide information about who is talking to whom about what (e.g., by way of deictic references - see [BAL], [VE1]).

Elements of Micro-level Awareness in Communilaboration

Gutwin and Greenberg [GUT] propose a framework for workspace awareness according to a number of elements that play a role in this form of awareness. For each element, they consider the mechanisms people use to gather awareness information. We have adapted their framework to include conversational awareness, adding the element *People* (refer to Table 1). We also defined the different elements in terms of their relation to the attentive states of others. We define an attentive state as a description of someone's focus of attention during an activity. At a syntactical level this involves describing the spatial and temporal properties of someone's (visual) attention, at a semantical level which actions, objects or people someone is attending to.

For each element of workspace awareness, Gutwin and Greenberg give their functionality by listing questions that participants might ask themselves during shared activities. In Table 1, we did the same for conversational awareness. Some elements have shared functionality between workspace and conversational awareness. These are represented in joint cells.

		Attentive State	Elements	Functionality	
				Workspace Awareness	Conversational Awareness
Syntax		Locus of Attention (*Spatial*)	Location	Where are they working?	Where are the people they communicate with?
		Attention Span (*Temporal*)	Presence	Who is participating?	
			Activity	How actively are they working?	How actively are they communicating?
Semantics	*Entity*	Attending to Objects Attending to People	Objects People	What object are they using or referring to? Whom do they work or communicate with?	
	Action	Attending to Actions	Action	What action are they performing or referring to?	
Pragmatics		Attention Range	Extents	What can they see?	What channels can they use?
			Abilities	What can they do?	Whom can they communicate with?
			Influence	Where can they make changes?	Where can they be?
		Future Attention	Intention (them)	What will they do next?	Whom will they communicate with next?
			Expectations (me)	What do they need me to do next?	Who wants to communicate with me next?

Table 1. *Organizing elements of micro-level awareness according to attentive state.*

Our model is hierarchically organized in three levels: the syntax, semantics and pragmatics of conveying awareness information in terms of attentive states. Each category of attentive state is attributed to one of these levels, and each element of awareness is attributed to a category of attentive state.

At the **syntax level** there are two categories, the basic building blocks of our model. *Locus of Attention* describes the spatial aspects of attention, while *Attention Span* describes the temporal aspects of attention. All higher-level categories in our model can be expressed in terms of these space/time coordinates. The next, **semantical level**, is functionally the most important. Users should always be aware what actions, objects and people other participants are attending to [RAE]. It is subdivided into **entity** and **action**. Entity identifies which objects or persons users are attending to at a given time. Action describes how this relationship varies over time. Thus, actions are described by the dynamics of attending to entities.

Categories at the **pragmatics level** heuristically describe expectations about the spatial and temporal behavior of others based on their history of attending to actions, objects and people. *Attention Range* relates to expectations in the spatial domain, while *Future Attention* relates to expectations in the temporal domain. Someone's *Attention Range* can be described by the spatial range of their history of attention to actions, objects and people, i.e., the space occupied by their behavior. Someone's *Future Attention* can be described by the rhythms of their behavior, based on a history of switching attention between actions, objects and people (turntaking behavior).

The present framework should be seen as an outline of a new design language for conveying awareness in groupware systems. Our syntax, semantics and pragmatics are levels of this language, not of the actual communication process. Such a language will also be of use in the analysis of existing task situations. By monitoring the participant's locus of attention—the syntax of our language—one can determine which objects (or other participants) they are attending to, in order to make higher-level inferences about the semantics and pragmatics of their (joint) activities, such as what actions they actually perform.

Conveying Awareness

Communicating awareness means that groupware systems should be able to collect awareness information on the input side and represent it on the output side.

Collecting Awareness Information

On the input side, the suggested framework for representing awareness as conveying the attention of others should make it easier for designers to systematically decide which input data to collect from participants. Much of this information can be collected in an implicit fashion and in terms of spatial and temporal measures: How long and where is someone looking?; How long and where has someone been moving his input device? An important consideration is that much of this information can be captured by monitoring existing input devices: mouse, cameras, microphone etc. An important new complement to such measures is the use of eye-movement information. Although at the moment such technology is not yet used for generic input purposes, this may well change in the near future. Capturing the actual focus and span of visual attention by means of an eyetracking system provides a relatively direct means of capturing awareness information about participants' relations to actions, objects and people [VE2]. Moreover, first experiments demonstrated that an explicit visualization of attentive states of partners improves *communilaboration* in constructive problem solving tasks involving experts and novices [VE1].

Visual Representation of Awareness Information

Since most people are experts when it comes to face-to-face communication, it seems reasonable to represent awareness using metaphors loosely based on face-to-face interaction. This way, the possibility of misinterpretations of these representations is minimized. Each element of awareness as listed in Table 1 should therefore have a representation with a meaningful correlation to a face-to-face situation. In a virtual meeting room, this might be accomplished as follows:

- **Conversational awareness.** Each participant can be represented by a *personification*: a functional model of a participant. The personification consists of a tile showing an image of the participant, which may be a photograph or a motion video image. A colored frame is used as a means of associating personifications with owned objects in a shared workspace. The orientation of personifications, placed within a 3D scene, can be used to convey the gaze direction of the participants in a meaningful way [VER].
- **Workspace awareness.** All participants' personifications are placed around a table in the 3D scene. This table represents a shared workspace on which they can place shared objects such as documents. Each participant's attention within this workspace can be represented by *lightspots* projected within the shared workspace according to the personification's orientation. Lightspots are associated with personifications by means of color coding. This "miner's helmet" metaphor can also be used to convey the locus of visual attention during document editing. When a shared document is opened, lightspots appear within the document, conveying where each participant is working. Temporal patterns and beam sizes directly afford awareness attributes such as level of activity and range of activity. Note that personifications can also contribute to workspace awareness since they can be rotated in such a way that they appear to look at a location on the table.

		Attentive State	Elements	Workspace Awareness	Conversational Awareness
Syntax		Locus of Attention (*Spatial*)	Location	Location of the lightspots on objects	Orientation of personification
		Attention Span (*Temporal*)	Presence Activity	Dynamics of the lightspots	Dynamics of orientation
Semantics	*Entity*	Attending to Objects	Objects	Position of objects; Position of lightspots on objects	Orientation towards objects
		Attending to People	People	Joint lightspot positions Joint orientation towards an object	Position of personifications; Orientation towards other personifications
	Action	Attending to Actions	Actions	Dynamics of attending to objects	Dynamics of attending to people
Pragmatics		Attention Range	Extents, Abilities & Influence	Spatial patterns in the dynamics of attending to objects	Spatial patterns in the dynamics of attending to people
		Future Attention	Intention & Expectations	Temporal patterns in the dynamics of attending to objects	Temporal patterns in the dynamics of attending to people

Table 2. *Representing elements of micro-level awareness according to attentive state.*

Table 2 shows how the suggested representations may provide answers to the questions in Table 1. The orientation of the personification and the location of the corresponding lightspot (i.e. the lightspot with the same color as the personification) convey the spatial aspects of someone's visual attention. From the movements of the personifications and the lightspots, people can see whether their partners are actually present, and if so, how actively they are working and communicating. These spatial and temporal aspects of awareness provide valuable cues for inferring attentive states at the semantical level. People working together on an object have their personification rotated towards the location of this object and their lightspots hovering around the object. When someone is speaking with other people, he will look at each of them from time to time [ARG], causing his personification to orient toward them. Actions can be inferred through the dynamic interactive behavior of lightspots, objects and personifications. Attention Range and Future Attention can be inferred through the spatial and temporal patterns found in a history of such behavior.

Recapitulating, we confined ourselves to representing explicitly only the spatial aspects of attentive states at the syntax and entity levels (at any given moment in time). All higher-level inferences about these representations are left to the user's interpretation. This does not mean that our framework would not allow explicit representation of higher-level attentive states. For example, one could implement Attention Range explicitly by translucently coloring parts of space where users have done things. However, by using attention-based metaphors modeled after everyday communication, we choose to structure the visual representation of awareness information in an implicit fashion, providing more or less natural affordances. Within conversational awareness, for example, gaze direction is used as a direct metaphor to convey interest, focus and intention during mediated communication. We experimented with its use in video mediated collaboration, and we demonstrated how still images conveying gaze direction improved conversational awareness with respect to full-motion video [VER; VO2]. Although our empirical findings are inconclusive in this respect, we strongly feel a representation of gaze direction can ease turntaking, particularly in large groups.

Applying the Framework: The GAZE Groupware System

With recent advances in hard- and software it has become possible to create multiplatform shared virtual meeting rooms supporting audio conferencing supplemented with micro-level awareness. We developed a prototype of such a system (The GAZE Groupware System) based on desk-mounted eyetracking technology and VRML 2.0. This *Virtual Reality Modeling Language* [SGI] allows interactive 3D scenes to be explored over the Internet with a standard multiplatform browser.

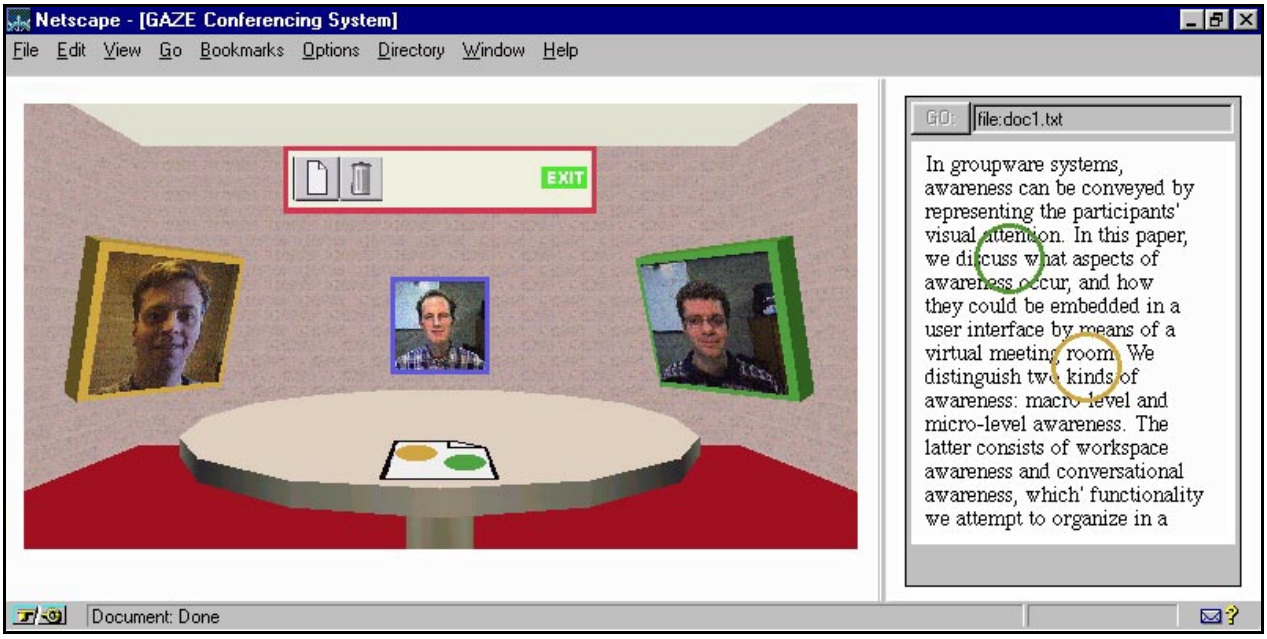

Figure 1. *The GAZE virtual meeting room.*

Figure 1 shows a typical participant's view during a four-person communilaboration using the GAZE Groupware System. The left part of this image contains a 3D scene showing a room with a table and the participants' personifications around it. Each personification consists of a simple 2D picture (or, in future versions, a live video image) which rotates in 3D space according to where their participant looks.

On the table, a document is placed with two lightspots on it. The lightspots belong to the persons sitting on the left- and right-hand side (and share the same color as their personifications[1]), indicating that their visual attention is confined to this document. Similarly, their personifications are tilted towards the document. When a participant opens the document on the table, it is downloaded and displayed in the right part of his screen. Here, lightspots indicate the location of other participants' visual attention *within* the document, providing a noncommand interface [NIE] which allows easy referencing of sections[2] ('What do you think of *this* bit?').

The system determines the location of the user's on-screen visual attention by means of an LC Technologies' *Eyegaze* eyetracker [LCT]. This way, the user's locus of attention is known, and can be displayed on the other participants' screens. Although current desk-mounted eyetracking technology

---

[1] Our coding scheme should be a redundant one. We might have gotten away with it had this publication appeared in color.

[2] We realize there are privacy concerns. At the moment, our only solution is to give continuous feedback about one's own lightspot, so that users are aware their point of gaze is being transmitted. We are not sure this is an appropriate solution.

still puts some restraints on the participant's head movements, we strongly feel that current developments are leading towards eyetracking technology which is inexpensive and totally transparent in use. As we are coming closer to understanding the relationship between dynamics of gaze behavior and the ongoing distribution of attention in its different forms [VE3], we can open the way to a whole generation of attention-based technologies.

Informal sessions with several hundred novice users at ACM Expo 1997 indicated that our approach to awareness representation in a mediated system seems to be a promising one. Most participants seemed to easily interpret the awareness information provided in terms of attention-based metaphors. The underlying eyetracking technology was, in many cases, *completely* transparent to the participants. We were surprised ourselves by the powerful presence effect generated by the rotation of personifications according to the locus of visual attention. More interestingly, this effect was achieved without the use of live video images (see [VER; VO2] for a more complete discussion). This is a very crude example of how attention-based internet services could actually lead to more optimal use of available network resources. In the foreseeable future, flexible use of bandwidth based on heuristics of the visual attention of individual users will become a reality [VO1].

Conclusions

In this paper, we have shown how many of the interpersonal awareness features in synchronous interactive communilaboration, particularly those on a micro-level, can be described in terms of our attentive state model. Our model allows groupware designers to conceptualize in a more structured way the kinds of awareness features they need to convey. It provides a way of thinking about capturing awareness information using direct yet transparent means and representing it across modalities using attention-based affordances. Our model is by no means exhaustive or complete. We consider it a simple reference framework which can be applied to a wide variety of situated communilaboration. As for our application, the GAZE Groupware System, we demonstrated how our framework may lead to improved awareness features without requiring any explicit additional input from the participants. The system not only shows how careful modeling of awareness features might improve distributed communilaboration, but also how it could lead to a more efficient use of network resources. We feel attention-based groupware systems have the potential of becoming an important and generic awareness supplement to multiparty speech communication over telephone systems and internet alike.

Acknowledgements

We would like to thank everyone who contributed either directly or indirectly to the development of the GAZE Groupware system. In particular, we would like to thank Nancy and Dixon Cleveland and their team @ LC Technologies for their wonderful support, the members of the Ergonomics Department, especially Robert Slagter and Harro Vons for their development work, David Kasik @ Boeing Commercial Airplanes, Axel Mulder @ Simon Fraser University, Arjen de Vries for his patience with the development team using his computer all the time and Michele Mariani @ the Vrije Universiteit Amsterdam for proofreading. Another special word of thanks goes to Harro Vons for his valuable contributions to this paper.

References

[ARG] Argyle, M. and Ingham, R. Gaze, Mutual Gaze and Distance. *Semiotica* 6, 1972, pp. 32-49.

[BAL] Ballard, D.H., Hayhoe, M.M., Pook, P.K. & Rao, P.N. Deictic Codes for the Embodiment of Cognition. *Behavioral and Brain Sciences*. 1997 In press.

[DOU] Dourish, P. and Bly, S. Portholes: Supporting Awareness in a Distributed Work Group. Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems, 1992, pp. 541-547.

[GAV] Gaver, W. Sound Support for Collaboration. In *Proceedings of ECSCW'91*. Amsterdam: Kluwer, 1991.

[GRE] Greenberg, S. Peepholes: Low Cost Awareness of One's Community. In *Companion of ACM CHI'96 Conference on Human Factors in Computing Systems.* Vancouver, Canada: ACM, 1996, pp. 206-207

[GUT] Gutwin, C. and Greenberg, S., Workspace Awareness for Groupware, *Companion of ACM CHI'96 Conference on Human Factors in Computing Systems*, Vancouver, Canada: ACM, 1996, pp. 208-209.

[HOF] Hoffman, J. Visual Attention and Eye Movements. In Pashler, H. (Ed.), *Attention*. London: University College London Press, 1997 In press.

[LCT] LC Technologies Inc., The Eyegaze Communication System, LC Technologies, Inc. Fairfax, VA. *http://www.lctinc.com*

[LOR] Lorenz, Ch. In Two Minds: Real Versus 'Virtual' Co-Location. *Financial Times*, 1995, November 10th.

[NIE] Nielsen, J. Noncommand User Interfaces. *Communications of ACM*, 1993, 36(4), pp. 83-99.

[RAE] Raeithel, A. and Velichkovsky, B. Joint Attention and Co-Construction: New Ways to Foster User-Designer Collaboration. In Nardi, B.A. (Ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction.* Cambrigde, MA: MIT Press, 1996.

[SGI] Silicon Graphics. The Virtual Reality Modeling Language 2.0. *ISO/IEC DIS 14772-1 submission*, 4 April 1997. *http://vrml.sgi.com/moving-worlds/*

[VE1] Velichkovsky, B.M. Communicating Attention: Gaze Position Transfer in Cooperative Problem Solving. *Pragmatics and Cognition*, 1995, 3(2), 199-222.

[VE2] Velichkovsky, B.M. and Hansen, J.P. New Technological Windows into Mind: There is More in Eyes and Brains for Human Computer-Interaction. *Proceedings of ACM CHI'96 Conference on Human Factors in Computing Systems.* Vancouver, Canada: ACM, 1996, pp. 496-503.

[VE3] Velichkovsky, B.M., Sprenger, A. and Unema,.P. Towards Gaze-Mediated Interaction: Collecting Solutions of the 'Midas Touch Problem'. In S.Howard, J.Hammond & G.Lindgaard (Eds.), *Human-Computer Interaction: Interact'97* (Sydney, July 14-18th), London: Chapman & Hall, 1997.

[VER] Vertegaal, R. Conversational Awareness in Multiparty VMC. *Extended Abstracts of ACM CHI'97 Conference on Human Factors in Computing Systems.* Atlanta, GA: ACM, 1997.

[VO1] Vons, J.A. *Human Interaction through Video Mediated Systems*. Master Thesis. Ergonomics Dept., Twente University, The Netherlands, 1996.

[VO2] Vons, J.A., Vertegaal R., and Van der Veer, G.C. Mediating Human Interaction with Video: the Fallacy of the Talking Heads. *Multimedia Minded Conference*, Lanaken, Belgium, 1997.

About the Authors

Roel Vertegaal is a computer scientist specializing in multimodal input/output and computer supported collaborative work. He is a research assistant with the Ergonomics Department at Twente University, The Netherlands. Vertegaal began his career in music, having done a Bachelor of Music degree at Utrecht School of the Arts where he studied and taught Music Technology and worked at the Center for Knowledge Technology on perceptual querying of musical audio databases. After completing an M.Phil. in Computer Science on the use of input devices in musical audio browsing at the University of Bradford, UK, he is now doing a PhD in Cognitive Ergonomics, focusing on eye input for collaborative systems and the role of gaze direction in human communication.

Boris Velichkovsky graduated in Neuropsychology from Moscow State University where later on he founded the first department for cognitive science. After a professorship at the University of Toronto, he is now Professor and Head of the Unit of Applied Cognitive Research at the expanding Dresden University of Technology in Germany. He is author of more than 100 papers on cognitive processes, communication and eye movements as well as several books such as Knowledge and Activity (Berlin: VCH, 1988), Communicating Meaning (Hillsdale, NJ: Erlbaum, 1996) and Usability Engineering (Stuttgart: Teubner, 1997). Boris Velichkovsky has been invited to give keynote addresses to several international congresses and conferences including CHI'96.

Gerrit van der Veer graduated in Cognitive Psychology at the Vrije Universiteit Amsterdam, with a PhD on mental models of computer systems. He is currently a reader in Computer Science, specializing in Interactive Systems, at the Vrije Universiteit Amsterdam in The Netherlands. He is also head of the Ergonomics Department at Twente University, The Netherlands. At both institutes he teaches cognitive ergonomics and user interface design. His research focuses on systematic design methods and Groupware Task Analysis. He is actively involved in the SIGCHI organization as Vice Chair for Conferences and was Co-chair of INTERCHI'93.

Author's Adresses

Roel Vertegaal (and Gerrit van der Veer)
Ergonomics Department
University of Twente
The Netherlands

roel@acm.org

Boris Velichkovsky
Unit of Applied Cognitive Research
Dresden University of Technology
Germany

velich@psy1.psych.tu-dresden.de

Gerrit van der Veer
Computer Science Department
Vrije Universiteit Amsterdam
The Netherlands

gerrit@cs.vu.nl