Designing Attentive Interfaces

Roel Vertegaal

Human Media Lab, CISC
Queen's University
Kingston, ON K7L 3N6
Canada
roel@acm.org
www.hml.queensu.ca

ABSTRACT

In this paper, we propose a tentative framework for the classification of Attentive Interfaces, a new category of user interfaces. An Attentive Interface is a user interface that dynamically prioritizes the information it presents to its users, such that information processing resources of both user and system are optimally distributed across a set of tasks. The interface does this on the basis of knowledge — consisting of a combination of measures and models— of the past, present and future state of the user's attention, given the availability of system resources. We will show how the Attentive Interface provides a natural extension to the windowing paradigm found in Graphical User Interfaces. Our taxonomy of Attentive Interfaces allows us to identify classes of user interfaces that would benefit most from the ability to sense, model and optimize the user's attentive state. In particular, we show how systems that influence user workflow in concurrent task situations, such as those involved with management of multiparty communication, may benefit from such facilities.

Keywords

Attentive Interfaces, Nonverbal Computing, Eye Tracking, Attention, User Interfaces.

1. INTRODUCTION

In recent years, there has been a resurge of interest in the use of eye tracking systems for interactive purposes. However, it is easy to be fooled by the interactive power of eye tracking. When first encountering eye based interaction, most people are genuinely impressed with the almost magical window into the mind of the user that it seems to provide. There are two reasons why this belief may lead to subsequent disappointment. Firstly, although current eye tracking equipment is far superior to that used in the seventies and early eighties, it is by no means perfect. For example, there is still the tradeoff between the use of an obtrusive head-based system or a desk-based system with limited head movement. Such technical problems continue to limit the usefulness of eye tracking as a generic input. Secondly, there are real methodological problems regarding the interpretation of eye input for use in graphical user interfaces. One example, the "Midas Touch" problem, is observed in systems that use eye movements to directly control a mouse cursor [20]. When does the system decide that a user is interested in a visual object? Systems that implement dwell time for this purpose run the risk of disallowing visual scanning behavior, requiring users to control their eye movements for the purposes of output, rather than input. However, difficulties in the interpretation of visual interest remain even when systems use another input modality for signaling intent. Another classic methodological problem is exemplified by the application of eye movement recording in usability studies. Although eye fixations provide some of the best measures of visual interest, they do not provide a measure of cognitive interest. It is one thing to determine whether a user has observed certain visual information, but quite another to determine whether this information has in fact been processed or understood [19].

Some of our technological problems can and will be solved. However, we believe that our methodological issues point to a more fundamental problem: What is the nature of the input information conveyed by eye movements and to

what interactive functions can this information provide added value?

For now, it seems, the most successful applications have been those in which the information conveyed by eye movement data is a direct measure of the locus of the user's (visual) attention. We argue that in many of these applications, the function of this information is to optimize the attention of the user, the attention of other users, or the processing resources of the computing system itself. We believe such applications in fact constitute a new category of user interface: the Attentive Interface.

2. BACKGROUND

The rationale for designing of artifacts in such a way that they optimize the attention of their users is not new. It existed long before personal computers were born and was the backbone of the European abstract functional art movement throughout much of the 20th century. According to Bauhaus director and architect Ludwig Mies van der Rohe, tools and other functional artifacts, such as buildings, needed to be designed using only those materials necessary to keep the construction structurally sound [3]. There were three main arguments to his "Less is More" rationale: 1) relinquishing unnecessary embellishments would improve the communication of the artifact's functionality; 2) simplifying the design would ease automated manufacturing while improving structural soundness, thus cutting costs; 3) the focus on the essence rather than on the peripheral would increase the esthetic value of the artifact. Figure 1 visualizes the design space constituted by the above trade-offs between the ease of use, engineering requirements and esthetics of a design. What Mies van der Rohe tried to argue was that if the right design choices are made, the above parameters need not be counteracting. Instead, they may complement and strengthen each other, as visualized in Figure 1 by a folding of the design space.

It is *exactly* this principle of economy of resources that drives human attention. Although humans have vast mental resources, they are of course limited. Norman [13] argued that for this reason, humans require a management system that selects and filters only that information that is most relevant. There has been much controversy as to how this selection process might operate. Von Helmholtz [8], and later Broadbent [2], considered human attention mostly as a physical filter, that selects information before entry into the brain.
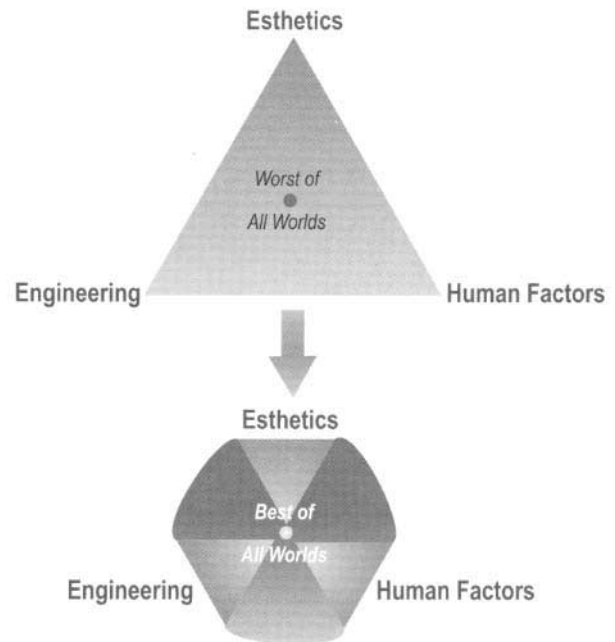
Figure 1. *Folding a design space such that constraints function as complements rather than tradeoffs.*

This is certainly the case with the eye, as the information bandwidth of the retina is a function of the distance to the center of the eye's visual axis. However, James [10], and later Deutch & Deutch [4], emphasized the (semantic) filtering that occurs after information has entered the brain. This is evidenced in the eye by the so-called Attentional Spotlight: that part of the retina that the brain selects for visual processing [14]. The controversy was not just a question of whether attentive filtering occurs early or late in human information processing. It was also a question of what determines the selection of information: is selection triggered by external stimuli or by anticipation through prior knowledge? This controversy was only recently concluded with Treisman's Feature Integration Theory: that attentive selection is guided by *both* input and cognition [17]. The question of input-driven versus knowledge-driven selection is very relevant to the design of an Attentive Interface. Many of the methodological problems associated with eye tracking applications are due to the fact that eye tracking input alone provides insufficient information for selective filtering.
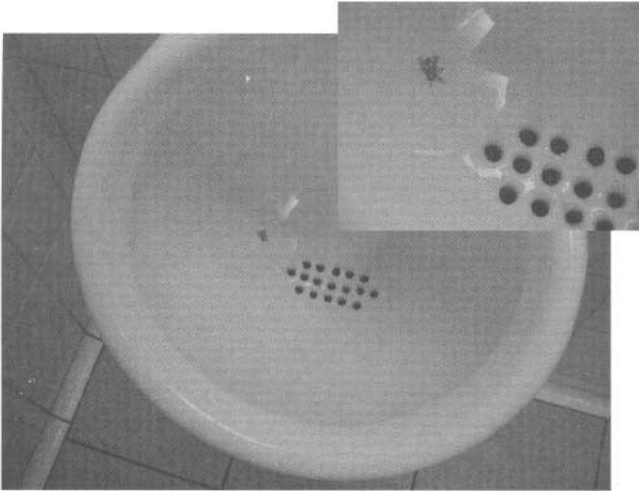
Figure 2. Passive attentive toilet design (Amsterdam Schiphol Airport, The Netherlands).

2.1 Early Attentive Interfaces

The question of input-driven versus knowledge-driven information filtering also relates Mies van der Rohe's design rationale to the economy of user information processing in the context of a set of tasks. We should design each tool such that the minimum information is provided to trigger the knowledge required to operate the tool [28]. We can thus interpret Mies van der Rohe's adage as a principle of information preservation. As any designer knows, this information comes in the form of the user's sequencing of intent and the task knowledge structures triggered by that intent [18]. An excellent example of how one can effectively influence the user's current intent on the basis of information about the user's goal intent is provided by the Attentive Toilets found at Amsterdam's Schiphol airport (see Figure 2). Indeed, the designer of these toilets has found a very simple yet effective mechanism to direct the intent of the user. However, in cases where the user's intent is in constant flux, artifacts need to have the capabilities to dynamically prioritize their information presentation on the basis of changes in that intent. Graphical User Interfaces were designed to do just that. According to Smith et al. [16], the rationale for the design of windowing systems is to allow users to focus on the task with the highest priority, in the context of other tasks of lower priority. In a typical graphical user interface, windows relevant to the present task will occupy the display space on which the user is currently fixating. Tasks of lower priorities are represented by icons that occupy peripheral vision. By selecting and positioning on-screen windows with a mouse, the user not only optimizes the information representation of the system according to her own attention, but also optimizes that of the computer system itself. By focusing processing priority on windows in the foreground, the resources of the system are optimally employed to comply with the current intent of the user. However, for the most part, it is the user's responsibility to optimize the economy of attention between system and user through a pointing device.

We do not have to look far to find examples of systems that seek information about the intent of their users through means that are much more proactive and implicit. A sophisticated application of this principle can be found in modern traffic light control systems. These actively seek information about the user's intent by implicitly sensing the focus of their attention through coils in the road surface. Traffic control systems do not rely solely on the input provided by their sensors. They apply knowledge rules based on statistical measures of traffic flow to decide which information to present to what user.

The above are all examples of interfaces predominantly occupied with the optimization of the user's current attention. A diary is a prime example of a tool that seeks to optimize the *future* attention of its user. As with traffic lights, the automation and interconnection of diaries, e.g., in hand held devices, has allowed the preemption of user attention through active interruption. However, as opposed to traffic lights, diaries know little about the current state of the user's attentive resources. The same applies to telecommunication and networked information systems such as email. It is in the notification structure associated with these systems that an attentive interface would perhaps be most pertinent. Indeed, according to Goldhaber [7], the Internet itself can be seen as an economy of the combined attentive resources of its users. He argues that in this economy the attention span of the consumers is a valuable limited resource. This is demonstrated by the current practice of internet based advertising agencies to sell the attention of web site users as measured by page views.

3. FEATURES OF ATTENTIVE INTERFACES

So what defines an attentive interface? An attentive interface is a user interface that dynamically prioritizes the information it presents to its users, such that information processing resources of both user and system are optimally distributed across a set of tasks. The interface does this on the basis of knowledge – consisting of a combination of measures and models – of the past, present and future state of the user's attention, taking into account the availability of system resources. As we have seen in the introduction, there can be many types of attentive interfaces. Given a set of examples, we will, however, try to identify a number of common themes, imposing the tentative classification shown in Table 1. Firstly, attentive interfaces can be differentiated on the basis of their ability to actively monitor the user's attentive state. The attentive toilet shown in Figure 1 is an example of a system that manipulates the attention of its user without any external measures. Although windowing systems of graphical user interfaces have no means to measure the user's attentive state implicitly, they do allow the user to communicate their attention explicitly through the manual organization of windows. We therefore classify these as *explicit*. Systems that employ sensing technology such as eye tracking to measure the user's current state of attention are classified as *implicit*. The next feature determines what is measured by such sensing technology: the locus of a user's attention, or its span. Typically, attentive interfaces will employ a combination of these techniques. Another important descriptor of an attentive interface is the type of sensing technology employed. In most explicit systems, this will simply be the user's pointing device. According to Vertegaal [26], systems that employ more implicit measures may use a combination of user presence, body orientation, head orientation and ultimately eye orientation to detect attentive state. Another important feature is whether the system stores any of the above measures as a model. For example, traffic control systems make their decisions on the basis of rules about the normal flow and volume of traffic. Finally, we consider the way in which an attentive system may increase or decrease the attentive load of the user, system, or network. A cell phone commands the user's immediate attention through active interruption, yet does little to optimize the user's mental load.

4. CLASSIFYING ATTENTIVE INTERFACES

However, the main classification provided by Table 1 is that according to utility: the task the Attentive Interface seeks to optimize. Our categories are ranked according to the level of mental processing required for the task.

Perhaps the most urgent category is that of attentive interfaces involved in the management of group communication. According to Duncan [6], turn taking in group conversations is very much a function of the optimization of group attention: listeners choose to be silent such that everyone can focus all attentive resources on a single speaker. Vertegaal et al. [24, 25] demonstrated that people indeed use information about each other attentive state, as given by their gaze direction, in this process. The GAZE Groupware System [21] is a video conferencing system that exploits this principle. It conveys the gaze direction of conversational partners across a network by rotating images of the participants toward the person they look at. The system measures whom participants look at using an eye tracker. Other user interfaces that fall in this category are avatar-based communication systems and multi-agent conversational systems such as FRED [23]. The next category deals with the management of offline communication. The sheer volume of today's email messaging has rendered the crude form of notification management present in today's email clients inappropriate. To solve this problem, Horvitz [9] developed *Priorities*, a messaging tool that collects attentive information – e.g., about the frequency with which users respond to messages from certain people – to prioritize the delivery of messages. Messages with a high priority rating are, for example, forwarded to an offline user's pager, while messages with low priority wait until the user checks them. Although *Priorities* builds a model of the user's attentive behavior, its ability to obtain information about the current state of the users' attention is limited. The third category deals with the opening and closing of dyadic communication channels. The EyeR glasses developed by Selker [15] use infrared light to sense whether a user is orienting his head towards another user. This information is then applied to regulate the opening and closing of communication channels, for example with a robot dog. When the user orients himself towards the robot dog, it starts barking. EyeContact, developed by Vertegaal [22] is a similar system which applies long range wearable

26

eye tracking technology to detect fixations at the user by other people. It can be used to modify the interruption state of incoming cell phone calls according to whether the user is currently engaged in a conversation, thus creating attentive cell phone technology [22].

Our next category pertains to what is perhaps the most basic feature of any attentive interface: the optimization of concurrent task execution. In the classic computer game Pong, the objective is to bounce a ball against a wall using a paddle. In traditional systems, the paddle is controlled by a joystick, leading to considerable eye-hand coordination problems. LC Technologies developed an eye-based pong that instead uses the horizontal coordinate of eye movements to control the position of the paddle, making this a game you cannot lose [11]. The windowing systems employed by graphical user interfaces are another example in this category. The most striking differences between Eye-based Pong and classic windowing systems is of course the absence of implicit measures in the latter. The MAGIC pointing system described by Zhai et al. [27] provides a hybrid between the two approaches. In MAGIC pointing, the user's eye coordinates are used to preposition the mouse cursor such that it reduces movement time toward a visual target.

Our final three categories pertain to the management of the user's basic attentive resources. Attentive interfaces involved in the management of the user's physical transportation seek to optimize the user's attentive resources at the presence level. As such, they regulate the physical movement of users. Traffic control systems are prime examples of this category. More basic systems in this category are automatic doors and other presence-operated devices.

One should note that attentive interfaces need not optimize the attentive resources of a user or system solely by a reduction in load. They can also empower the user by enlarging their attentive capacity. Indeed, most gaze contingent displays seek to augment the basic attentive resources of the user or system. Attention-based video compression [5] and gaze-contingent 3D level-of-detail rendering [12] are examples of this category. Our final category pertains those interfaces with the most basic of attentive tasks: to attract the user's attention. The attentive toilet shown in Figure 1 is a prime example of this, as are most security monitoring systems.

5. DISCUSSION

The proposed classification allows us to reason explicitly about the design attributes of attentive interfaces. It allows us to identify the categories of user interfaces that are most likely to require attentive enhancements through tracking technologies and priority models. We have listed a number of these candidates in Table 1, some of which we will now discuss. Likely candidates for attentive enhancement in the category of group communication systems are systems that represent users in a virtual space in the form of avatars. Some systems in this category already employ a crude form of attention management that uses co-alignment of avatars to decide whether to open or close audio connections [1]. However, systems in this category are very likely to benefit from the use of more implicit sensing technologies, such as the measurement of eye, head or body orientation. In the category of dyadic communication management, cell phones are the most obvious candidates for improvement. Interruption of multiparty conversation through cell phones has become such a problem that it has prompted the installation of cell phone jamming technology in certain meeting areas. Cell phones should be capable of adjusting their interruption priorities according to the user's current state of attention [22].

From our categorization it also becomes clear that there are very few systems indeed that employ a model of the user's past attention in their prioritization of current information. One reason for this is that the required statistical reasoning techniques are still in their infancy [9]. As Microsoft's Office Assistant exemplifies, it is better not to incorporate any model than it is to incorporate a poor model. However, we believe that as the technology improves, we will increasingly see statistical learning techniques incorporated in attentive interface design.

	Active	Locus/Span	Sensor	Attention Model	Active Interrupt	Reduce Load
Group Communication Management						
GAZE Video Conferencing	Implicit	Locus	Eye			User System Network
Attentive Agents	Implicit	Both	Eye		•	User System
Avatars	Explicit	Locus	Hand			User Network
Offline Communication Management						
Priorities System	Implicit	Span	Hand	•	•	User Network
Email	Explicit	Span	Hand		•	None
Diary	No					User
Headline	No					User
Dyadic Communication Management						
EyeR	Implicit	Both	Head			User System network
EyeContact Sensor	Implicit	Both	Eye			User System Network
Cellphone	Explicit	Span	Hand		•	None
Concurrent Task Management						
Eye-based Pong	Implicit	Locus				User
MAGIC pointing	Implicit	Both	Eye			User
Windows	Explicit	Both	Hand		•	User System
Physical Transportation Management						
Presence Operated Devices	Implicit	Both	Body		•	None
Traffic Control Systems	Implicit	Both	Body	•	•	User System
Door	No					None
Augment Attentive Resources						
Gaze-Contingent Displays	Implicit	Locus	Eye			User System
Attention-based Compression	Implicit	Both	Eye			User System Network
Canon EOS-5	Implicit	Locus	Eye			User System
Direct Attentive Resources						
Security Monitoring Systems	Implicit	Locus	Body		•	User
Attentive Toilet	No					System

Table 1. Classification of Attentive Interfaces with examples.

6. CONCLUSIONS

In this paper, we proposed a tentative framework for the classification of Attentive Interfaces, a new category of user interfaces. We defined an Attentive Interface as a user interface that dynamically prioritizes the information it presents to its users, such that information processing resources of both user and system are optimally distributed across a set of tasks. The interface does this on the basis of knowledge – consisting of a combination of measures and models – of the past, present and future state of the user's attention, given the availability of system resources. Measures are provided by tracking presence, body orientation, head orientation or eye orientation of users. Models may consist of rules and heuristics based on previously observed patterns of attentive behavior. We have shown how Attentive Interfaces provide a natural extension to the windowing paradigm of the Graphical User Interface. One of the main benefits of our classification of Attentive Interfaces is that it makes explicit those categories of user interfaces that would benefit most from the ability to sense and model the user's attentive state. In particular, we showed how systems that influence user workflow in concurrent task situations, such as those involved with management of multiparty communication, may benefit from such facilities.

7. ACKNOWLEDGEMENTS

8. REFERENCES

1. Benford, S., Greenhalgh, C., Bowers, J., Snowdon, S., and Fahlén, L. User Embodiment in Collaborative Virtual Environments. In Proceedings of CHI'95. Denver, Colorado: ACM, 1995.

2. Broadbent, D.E. Perception and Communication. London: Pergamon Press, 1958.

3. Carter, P. and Mies van der Rohe, L. Mies Van Der Rohe at Work. Phaidon Press, 1999.

4. Deutsch, J.A. and Deutsch, D. Attention: Some Theoretical Considerations. Psychological Review 70, 1963, pp. 80-90.

5. Duchowski, A.T. and McCormick, B.H. Gaze-Contingent Video Resolution Degradation. In Human Vision and Electronic Imaging III. San Jose, CA USA: SPIE, 1998.

6. Duncan, S. Some Signals and Rules for Taking Speaking Turns in Conversations. Journal of Personal and Social Psychology 23, 1972.

7. Goldhaber, M. The Attention Economy and the Net. First Monday http://www.firstmonday.dk, 1997.

8. H. Helmholtz, von. Helmholtz's Treatise on Physiological Optics. Southall, J.P.C. (Ed.). New York: Dover Publications, 1962.

9. Horvitz, E., Jacobs, A., and Hovel, D. Attention-Sensitive Alerting. In Proceedings of UAI '99, Conference on Uncertainty and Artificial Intelligence, Stockholm, Sweden. Morgan Kaufmann: San Francisco, 1999, pp. 305-313.

10. James, W. The Principles of Psychology. New York: Holt, 1890.

11. LC Technologies, Inc. The Eyegaze Communication System. Fairfax, VA: http://www.eyegaze.com, 1997.

12. Murphy, H. and Duchowski, A. Gaze-Contingent Level of Detail Rendering. In EuroGraphics 2001. Manchester, UK, 2001.

13. Norman, D. Towards a Theory of Memory and Attention. Psychological Review 88, 1968, pp. 1-15.

14. Posner, M.I. Orienting of Attention. Quarterly Journal of Experimental Psychology 32, 1980, pp. 3-25.

15. Selker, T. Eye-R, a Glasses-Mounted Eye Motion Detection Interface. In Extended Abstracts of CHI'01. Seattle, WA: ACM, 2001.

16. Smith, D., Irby, C., Kimball, R., Verplank, B., and Harslem, E. Designing the Star User Interface. Byte 7(4), 1982, pp. 242-282.

17. Treismann, A. and Gelade, G. A Feature-Integration Theory of Attention. Cognitive Psychology 12, 1980, pp. 97-136.

18. Van der Veer, G.C., Yap, F., Broos, D., Donau, K., and Fokke, M.J. ETAG - Some Applications of a Formal Representation of the User Interface. In Proceedings of INTERACT'90. Amsterdam: North-Holland, 1990.

19. Velichkovsky, B. and Hansen, J.P. New Technological Windows to Mind: There is More in Eyes and Brains for Human-Computer Interaction. In Proceedings of CHI'96. Vancouver, Canada: ACM, 1996, pp. 496-503.

20. Velichkovsky, B., Sprenger, A., and Unema, P. Towards Gaze-mediated Interaction: Collecting Solutions of the "Midas Touch Problem". In Proceedings of INTERACT '97. Sydney, Australia, 1997.

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

22. Vertegaal, R., Dickie, C., Sohn, C. and Flickner, M. Designing Attentive Cell Phones Using Wearable EyeContact Sensors. Submitted to Extended Abstracts of CHI'02. Minneapolis: ACM, 2002.

23. Vertegaal, R., Slagter, R., Van der Veer, G.C., and Nijholt, A. Why Conversational Agents Should Catch the Eye. In Extended Abstracts of CHI'2000. The Hague, The Netherlands: ACM, 2000.

24. Vertegaal, R., Slagter, R., Van der Veer, G.C., and Nijholt, A. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In Proceedings of CHI'01. Seattle, WA: ACM, 2001.

25. Vertegaal, R., Van der Veer, G.C., and Vons, H. Effects of Gaze on Multiparty Mediated Communication. In Proceedings of Graphics Interface 2000. Montreal, Canada: Morgan Kaufmann Publishers, 2000, pp. 95-102.

26. Vertegaal, R., Velichkovsky, B., and Van der Veer, G.C. Catching the Eye: Management of Joint Attention in Cooperative Work. SIGCHI Bulletin 29(4), 1997.

27. Zhai, S., Morimoto, C., and Ihde, S. Manual and Gaze Input Cascaded (MAGIC) Pointing. In Proceedings of CHI'99. Pittsburgh, PA: ACM, 1999, pp. 246-253.

28. Zhang, J. and Norman, D. Representations in Distributed Cognitive Tasks. Cognitive Science 18, 1994, pp. 87-122.