AURAMIRROR, REFLECTIONS ON ATTENTION:

Visualizing Attention in Interpersonal Communication

by

ALEXANDER WILLIAM VISCHER SKABURSKIS

A thesis submitted to the School of Computing
in conformity with the requirements for
the degree Master of Science

Queen's University
Kingston, Ontario, Canada
January, 2004

Your file Votre référence
ISBN: 0-612-92353-3
Our file Notre référence
ISBN: 0-612-92353-3

Abstract

This thesis examines the ways in which computers can recognize, monitor and display the foci of attention among people within a defined *attentive space*. It develops the prototype system *Auramirror* that can record the negotiation of attention and display it to the participants to gain the feedback of participants or researchers when evaluating attention-tracking algorithms. The prototype was deployed at the Ontario Science Centre in Toronto where over fifteen thousand people participated in the exhibit over the past four months. A user evaluation was carried out using descriptive survey research techniques.

The thesis places the problem of attention recognition in the context of Human-Computer Interaction and discusses some of the precedents of interfaces that use attention to improve interaction. This work also draws on examples of the use of art to inform the design process and broaden the possibilities for the links between people and machines. It sets the background for the development of the prototype mirror that reflects the users and superimposes on their images *attention auras* representing potential attention as well as *attention tunnels* representing mutual attention between individuals.

The prototype was developed using Intel's open source image processing library (OpenCV), and the DirectShow Software Development Kit (SDK) by modifying and integrating a number of algorithms from these sources as well as creating some new algorithms. A boosted classifier was used to find faces in a captured image from the system's video camera and a Kalman filter was used to track the found faces. The auras, which represent potential attention, were modeled as an elastic, spherical, triangular mesh that was created as an icosahedron-based geodesic sphere.

The system was deployed at the Ontario Science Centre in Toronto and may become a permanent exhibit demonstrating computer perception. Over a one month period, 121 participants completed the questionnaire. Their assessment suggests that Auramirror does accurately estimate attention shared between pairs of people for a limited context. Suggestions for improvements and enhancements complete this thesis.

Acknowledgements

I would like to thank my thesis supervisor Dr Roel Vertegaal for his guidance, his support and his friendship. Without his participation, this thesis and the Auramirror prototype could not have been developed. I appreciate greatly the help that Dr Henk Meijer gave in reviewing and encouraging my progress as a student and in commenting on the drafts of this thesis. I would like to thank Ms Ana Klasnja of the Ontario Science Centre for help in setting up the evaluation of Auramirror. I would also like to express my gratitude to my parents Andrejs Skaburskis, Sandra Olney, Jacqueline Vischer and John Zeisel for sharing their views on the purpose and potential of graduate studies, for editorial help, and for their patience, persistence and expertise in raising me. Finally, I would like to recognize the support of all my friends in Dr Vertegaal's Human Media Lab, particularly Jeffery Shell for his help in developing the concepts behind this research, and Connor Dickie for his inspiration on the title of this thesis. I would finally like to thank Josephine Chu for her moral support and camaraderie. To myself I credit all errors herein.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Acronyms and Abbreviations

AUI .... Attentive User Interface
COM .... Component Object Model
GLUT .... Open source Graphics Language Utility Toolkit
MFC .... Microsoft Foundation Classes
OpenCV .... Open Computer Vision Library
PDA .... Personal Digital Assistant
HCI .... Human-Computer Interaction
HSV .... Hue, Saturation and Value
RGB .... Red, Green and Blue
SDK .... Software Development Kit
UC .... Ubiquitous Computing

1
INTRODUCTION

This thesis considers issues within the field of Human-Computer Interaction (HCI). It does so in the context of developing a system called Auramirror for recognizing, tracking, recording and displaying the attention people pay to each other during conversations when they are within an *attentive space*. This effort hopes to advance the field of HCI by trying to improve the interaction between people and computers so that the whole human-computer system is more efficient and useful.

The importance of improving human-computer interaction increases as the standard configuration of a single user using a single computer workstation is changing into a single user using many computers. It is not uncommon for individuals now to own and operate a desktop computer, a laptop computer, a personal digital assistant, and a mobile phone. As networks proliferate, the typical paradigm of computer use may become many users using many computers. Under these new circumstances, many existing user interfaces may break down and lose functionality. These changes in the roles and pervasiveness of technology challenge designers and researchers to create new methods and rules to update the way that humans and computers interact.

1.1 CONTRIBUTION

The work reported in this thesis contributes to the field of HCI in three ways: it proposes a specific problem of human-computer interaction; it develops a prototype system for dealing with that problem; and it demonstrates and evaluates the prototype in actual use for its ability to visualize attention. The specific problem identified addresses the ways in which computers can passively measure and visualize data on the attention people pay to each other during face-to-face interactions. To address this problem, a system, Auramirror, was developed to use computer vision to track the attention shared among participants who enter within the view of the system's video camera. A prototype was produced and evaluated at the Ontario Science Centre.

Auramirror presents a novel way of visualizing the attention the participants pay to each other in an attentive space. This work synthesizes existing computer vision algorithms into a process that estimates and tracks attention in real-time. The design of the system was conceived to allow a live presentation of the visualization to an audience in order to promote an awareness of attention interactions among audience members, and to create the ability to determine the quality of the estimates made by the system. The system was also designed to be artistic so that the audience's awareness of attention and how it is manipulated by technology can be addressed by an artistic dialog between work and audience. A refereed paper on the Auramirror system has been published [51] in the proceedings of SIGCHI Conference on Human Factors in Computing Systems, reputed to be "the leading international forum for the exchange of ideas and information about computer-human interaction" [9]. Another will be published in the proceedings of ETRA, Eye Tracking Research and Application Symposium 2004 [52]. The system is currently on exhibit at the Ontario Science Centre, which has expressed an interest in it as a permanent exhibit on computer perception.

1.2 OUTLINE

The thesis is presented in six chapters. Chapter 2 gives a brief history of the field of HCI from its pre-inception roots up to the recent topics upon which this thesis focuses. It will define usability, a central theme of HCI research, and detail the specific problem that the thesis proposes. Chapter 3 presents a detailed description of Auramirror, the system designed and implemented for this thesis in response to the challenge expressed in Chapter 2. Chapter 4 reports the specifics of Auramirror's implementation. It discusses the hardware requirements, programming infrastructure, new algorithms used in the graphical visualization of attention, existing computer vision algorithms used and modifications on these, and how all these algorithms are synthesized into the estimation, tracking and display of attention. Chapter 5 describes a survey of users conducted with the expectation of justifying the value of the visualization technique used by Auramirror and validating some claims about how Auramirror is perceived by its audience. Chapter 6 suggests some directions for future research and states the conclusions of the research.

2
CONTEXT

Within computer science, the field of HCI studies the way people use computers and explores the means by which computers can be made more usable. This is accomplished through the study and design of the *user interface*, the boundary between the computer system and the person or people using it. This chapter describes and discusses some of the major advances in the field of HCI. It begins with the early work by Vannevar Bush that, even before the creation of HCI as a field, established the overarching motivations and themes of this field. Current modes of thought in the areas of Ubiquitous Computing and Attentive User Interfaces are described, and specific pursuits of contemporary research are referenced. This is followed by a definition of *usability* and an explanation as to why usability is important to computing. Finally, the specific problem that will be addressed in this thesis is described.

2.1 PEOPLE-COMPUTER SYMBIOSIS

In 1945 Vannevar Bush [6] laid the foundations for the philosophy that directs HCI research when he proposed the application of science to expand the human capacity for thought. Scientific research had so far been applied to extending the physical capacities of humans, making them faster and stronger, but, at the time of his seminal article, scientists needed new directions to explore. Vannevar Bush identified the problem created by our limitations in making use of the vast quantities of materials that were being published.

> The summation of human experience and knowledge is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same that was used in the days of square-rigged ships [6].

Vannevar Bush proposed *memex* as an engineering solution to illustrate the power of cooperation between humans and machines. Memex would store vast quantities of personal writing and published literature and then index it for easy access and make it readily available for perusal on a desktop viewing panel. Furthermore, to help establish associative connections, Bush proposed that memex be able to save *trails*: information about the relations identified between items and used to traverse a collection of literary sources much like a sequence of links between documents on today's World Wide Web. He believed that the ability to share both the collection of literature and the trails the user may have created when going through the literature would allow for a more versatile and efficient way for researchers to collaborate and find the material they need.

Bush's foundation for HCI paralleled the theme and quest of the artists founding the Futurist movement between 1909-1944. Futurism revelled in a more extreme celebration of machines. Indeed the Futurists aimed to provoke strong reactions from the public with their many manifestos. Futurism constituted "the first 'modern' attempt to reorganize art and society around technology and the machine ethic" [41]. Marinetti, the creator of Italian Futurism, proclaimed, "we will conquer the seemingly unconquerable hostility that separates our human flesh from the metal of motors." [34 via 41] Indeed, this is a bold statement that nonetheless foreshadows the aims of HCI, but through art.

J.C.R. Licklider advanced the work on HCI in the 1960s when he introduced the term *man-computer symbiosis* [30]. He envisioned a partnership whereby the respective strengths of computer systems and human cognition would enable the combination of humans and machines to "think as no human brain has ever thought and process data in a way not approached by the information handling machines we know today" [30]. Douglas Engelbart complemented Licklider's vision; he wrote about augmenting our intellect by using computers to improve our ability to find solutions to complex problems [13]. Douglas Engelbart's contributions to HCI were formidable. He invented the mouse, and his 1968 presentation of work originally inspired by memex was a major step towards the development of one of the most significant advances in the field of HCI: the graphical user interface [25]. The visions of Bush, Licklider and Engelbart formulated the basic challenge of the interface between humans and machines that is still relevant today. They identified, in some measure, that the real breakthrough of the computer lies in the idea of it as

> a symbolic system, a machine that traffics in representations or signs rather than in the mechanical cause-and-effect of the cotton gin or the automobile [25].

The field of HCI aims to increase the utility of computers by studying the interface to this abstraction that is the modern computer.[1]

[1] Licklider also foresaw the utility of computers specifically in communication, "we are entering a technological age in which we will be able to interact with the richness of living information—not merely in the passive way that we have become accustomed to using books and libraries." [31] Licklider goes on to introduce the notion of on-line communities connected through a *supercommunity* of networked computers. In some ways Auramirror is an attempt to visualizing this connection between humans, with and through computers, in the context of communication.

2.2 UBIQUITOUS COMPUTING

Ubiquitous Computing (UC) [61] is Mark Weiser's name for his proposed model of the interaction between computers and humans. Mark Weiser argued that a good tool is one that is known so well that it effectively becomes invisible in use. When technologies permeate our environment to a sufficient degree they can become invisible. He argues that computing has the potential to join the ranks of other ubiquitous technologies such as written language, electricity and plumbing. He posits that processing power will become the invisible infrastructure of the future that is used by all manner of devices. He envisions the deployment of large numbers of different types of computers designed to enhance the environment. These computers could be interconnected through a ubiquitous wireless network, and they would be aware of their own location as well as the presence, and possibly the identity, of collocated people. This will allow computers to perform helpful tasks and store useful data, based on their context and without the need for sophisticated artificial intelligence systems, whose development would pose considerable challenges in their own right.

2.2.1 Calm Technology

The next step towards the development of UC, Weiser suggests, is the design of *calm technology* [62]. Calm technology is characterised by the use of peripheral channels to convey information, and by its ability to move smoothly in and out of the user's focus. Users can vary their degree of awareness of the information being presented to them by shifting focus to peripheral sources at their convenience. The example Weiser gives is that of a newspaper: a substantial source of information easily put aside and moved into the reader's periphery. Nevertheless it is possible that a headline catches the reader's eye, and his or her attention shifts focus to the newspaper article, which can then be read. This principle can be applied to technology to make abundant information available, and interruptions potentially less intrusive, by using subtler notification methods and peripheral channels. Calm technology allows the increase of immediately available information without overly increasing the mental effort required by the user to access this information.

A key aspect to calm technology is its ability to move smoothly between the periphery and the focus of the user's attention. An example of UC and calm technology in use is found in *informative art*. Informative art objects are designed to convey abstractions of information in an unobtrusive and aesthetic way [44]. Monitors hang on the wall as dynamic paintings. The colours and patterns in the painting are computed based on a mapping of current data, such as the time of day or the number of e-mails being transmitted through the local server. However, the principle goal of informative art is not to visualize data but to promote reflection and concentration. Informative art offers a justification and design criteria for demonstrating information to users in a way that is primarily aesthetic.

2.2.2 Context-Aware Computing

Current desktop computers have little awareness of their user aside from mouse movement and keyboard presses. UC, and, more specifically, the *context-aware* components described by Gregory Abowd address this shortcoming [1]. Processing power has increased to the point that it is now possible for computers to use input means beyond the usual mouse and keyboard, not only for the user to communicate more thoroughly, but to potentially remove the need for explicit user input altogether. The computer can be developed to make use of data it gathers about its own context. The context-awareness theme of UC focuses research on the ways to measure and use contextual data. Abowd defines context-aware computing as

> any attempt to use knowledge of a user's physical, social, informational and even emotional state as input to adapt the behavior of one or more computational services. [1]

This applies to mobile devices that can make use of information such as location, and also applies to computer-enhanced environments in which computers recognize those who are present and what they are doing.

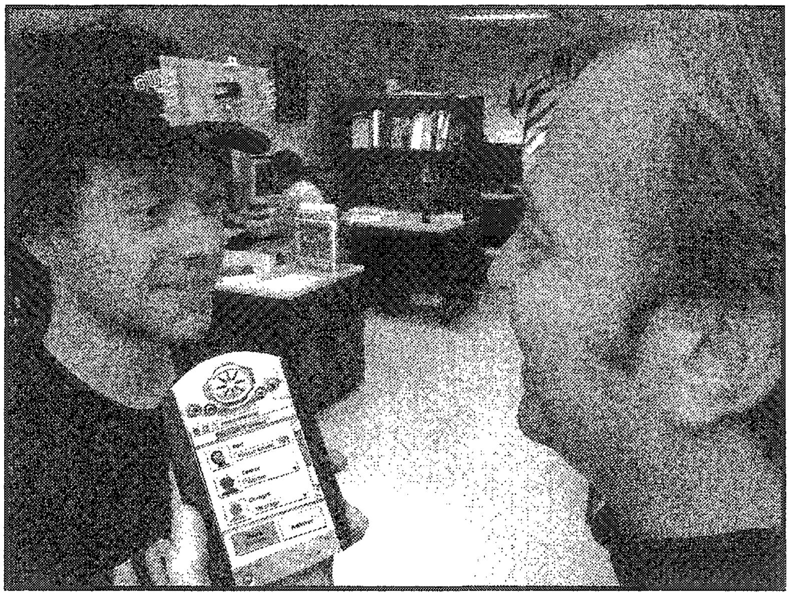

Figure 1. Attentive Cell Phone with Eye-Contact Sensor. [56]

For example, Vertegaal's attentive mobile phone collects contextual information indicating whether or not a user is involved in a face-to-face conversation [56]. He mounts an eye-contact sensor on the user's head and orients it so it will detect eye contact from people in the user's field of view (Figure 1). The user wears a hands-free phone attachment consisting of an earpiece and microphone combination. Even when the phone is not in use, the system collects data on the user's speech activity sensed by the microphone. Detection of eye contact, combined with the information on speech activity, is used to estimate whether the user is currently in a face-to-face conversation. The status of the user is transmitted to other people on the user's contact list. An instant messenger-style status display on a wirelessly connected personal data assistant (PDA) informs a potential contact of the user's state as illustrated in Figure 1. Instead of having a system that automatically filters interruptions, Vertegaal's system allows the potential callers to make informed decisions about the current availability of the person they are thinking about interrupting. Furthermore, if the caller decides to interrupt, he or she can use the information gained through the system to help decide whether to interrupt the person publicly with the usual ring or privately with an in-ear knocking through the wearer's headphones. This example illustrates how simple contextual data can be gathered and used to inform negotiation of technology-mediated interactions. It also demonstrates how peripheral channels can be used to facilitate this negotiation.

Instead of a plethora of mobile computers, UC theory and context-aware computing can be applied to environments by embedding sensors in a space to make it a *perceptive space*. User can then make use of the perceptive abilities of the space in the same way that they rely on the perceptive abilities of other people. These perceptive spaces, of which Wren's *Smart Room* [66] is an example, should not require the users to wear specialized hardware for input. This room contains a large projection screen and several video cameras and microphones. By using computer vision and AI techniques, the Smart Room can gather and respond to user input such as speech or gestures. By allowing users to engage the system without becoming separated from their surroundings provides for a more natural interaction with the system. Opening and closing dialogues with the system is fast and easy, and the user is not isolated from other users or objects in the room. Perceptive spaces are an alternative to virtual reality systems. Instead of trying to replace reality, perceptive systems aim to integrate computers into everyday life. Although the smart rooms developed by Wren described here entail, for the most part, a single user and single computer paradigm.

Trevor Darrell developed a virtual mirror [11], a perceptive system with technical requirements similar to the Auramirror system that will be described later. A virtual mirror used real-time face tracking to produce a system that would react to its audience. In this implementation, a face is chosen from the audience and graphically warped by the computer. The comically modified mirror video image is displayed as a form of entertainment. Users of the system reported that they felt that the system “knew” where their faces were.

2.3 ATTENTIVE USER INTERFACES

Vertegaal’s attentive mobile phone is an example of an Attentive User Interfaces (AUI). AUIs are an extension of context-aware and UC that attempts to solve the problems that arise when multiple computers vie for the attention of one or more users [48]. In multi-party, face-to-face communication, humans manage each other’s attention, negotiating how its focus changes so that messages are delivered efficiently and so that the disruptiveness of interruptions is minimized. Computers might become more sociable if they could effectively mimic humans in this respect.

2.3.1 Significance of Gaze

The development of AUIs was based in part on social psychological theories about the regulation of human multiparty communication. People can only absorb the message from one single verbal stream at a time and have developed ways to focus and manage attention to compensate for this limitation [48]. One such solution for managing group conversations is *turn-taking*: a means by which speakers in a conversation alternate so that only one person speaks at a time. According to Short, as many as eight cues may be used to negotiate the exchange of speakers [50]. In group conversations of more than two people, only eye contact indicates to whom the speaker may be yielding the floor [2].

Eye contact is a nonverbal visual signal that can be used to negotiate turns without interrupting the verbal auditory channel. Gaze implicitly indicates the object of a person's attention [24]. This is one way that humans use multiple channels to improve the efficiency of interpersonal communication.

In addition to its role as a selective mechanism, gaze plays a role in encouraging communication. People are more likely to speak in multiparty conversation when they are subject to increased eye contact [59]. Gaze may induce arousal in the subject and so may encourage responses from recipients of the gaze. Argyle has shown that the amount of eye-contact between two people relates to the intimacy felt between them [3]. Computer systems aware of a user's gaze may be able to make use of some of the information transmitted by patterns of gaze.

Psychology has explored the association between gaze and attention. For this thesis, *attention* is defined roughly as the physiological and cognitive filters that one applies to sensory input. This definition is in close accordance to Kahneman's definition of selective attention as the internal mechanism that determines the significance of some stimuli in preference to others that will be allowed, in turn, to control behaviour [26].

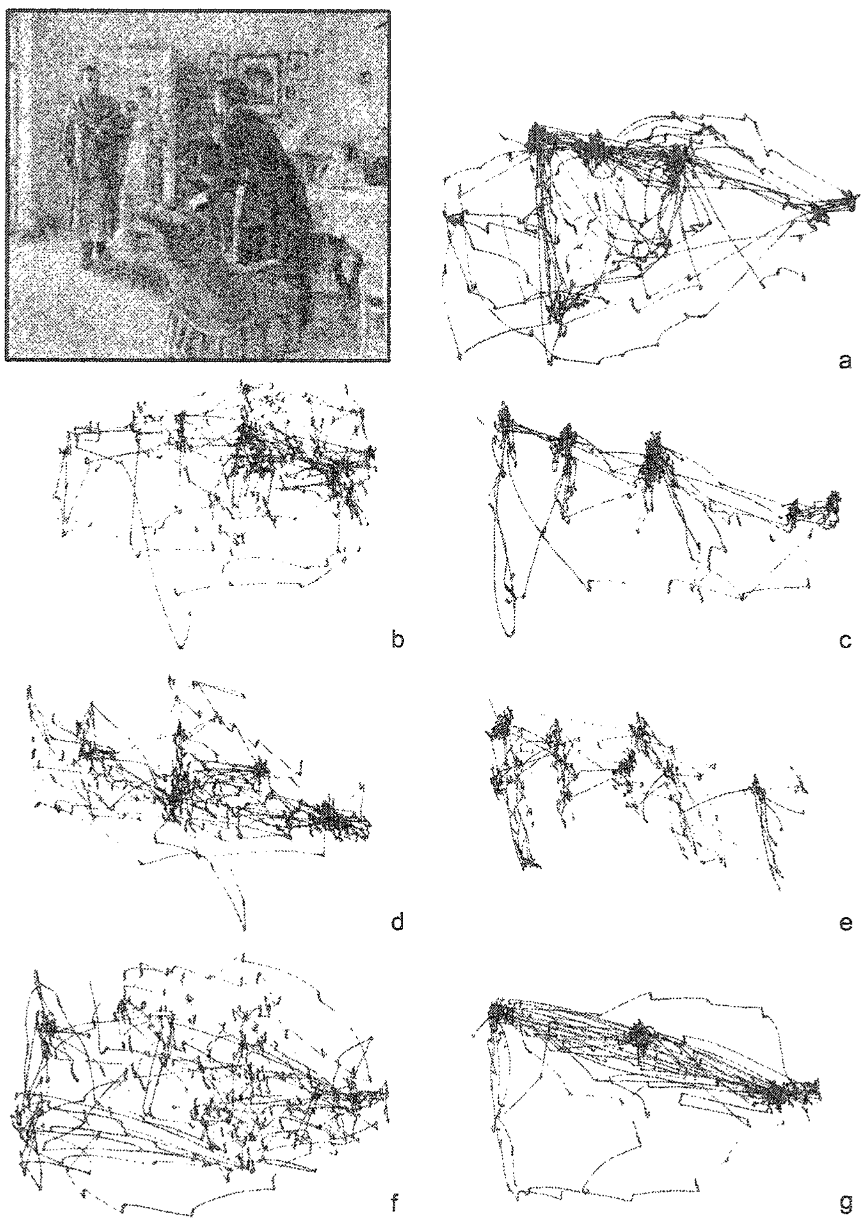

Figure 2. Seven records of eye movements by the same subject. Each record lasted 3 minutes. a) Free examination. Before subsequent recordings, the subject was asked to: b) estimate the material circumstances of the family; c) give the ages of the people; d) surmise what the family had been doing before the arrival of the "unexpected visitor;" e) remember the clothes worn by the people; f) remember the position of the people and objects in the room; g) estimate how long the "unexpected visitor" had been away from the family [68].

To support the association of interest and attention Yarbus provides evidence that people look at what they are interested in [68]. He asked subjects to ascertain specific aspects of a scene depicted by a picture. The eye movements recorded during the subject's subsequent examination of the image were determined by the information the subject was trying to discover (see Figure 2). Current models of visual attention describe gaze as indicating the focus selected by pre-attentive processing [54,20]. Starker and Bolt used this association between interest and gaze to develop a system that used gaze to infer the interest a user had in displayed objects. Based on Antoine de Saint Exupery's book, *The Little Prince*, 3-D graphics representing the prince's world was displayed to subjects whose gaze was measured by an eye-tracker. If subjects looked generally at the scene an overall narration was heard. If the user looked at a single feature of the world or at a group of features then the narrator gave appropriately specific commentary. The system shows how gaze patterns can be used in real-time to inform the system so it produces the correct narrative.

Information on the user's gaze is an ideal candidate for computing devices that try to sense their user's attention. Gaze information can be collected in a separate input channel to allow computers to determine whether a person is attending to them or to some other device or person. By tracking whether a user ignores or accepts requests for attention, interruptions by devices can be designed to be more subtle and sociable. Gaze is relatively easy to measure with eye-tracking methods, eye-contact detection devices or computer vision techniques that record head orientation. Studying attention by monitoring a person's gaze makes the identification and measurement of the focus of attention more manageable both computationally and perceptively.

2.3.2 Principles of Attentive User Interfaces

Methods for identifying attention can be used to improve ubiquitous user interfaces by increasing their ability to imitate inter-personal communication when initiating dialogue with users. The twin problems of multi-party communication and coordinating interruptions by ubiquitous devices are very similar. In both situations there is competition for the limited resource of human attention in order to deliver a message. However, people know when and how to interrupt each other in such a way that the interruption is least disruptive and most appropriate. Turn-taking provides a powerful metaphor for the regulation of communication between people and ubiquitous devices. By tracking whom or what their user is paying attention to, computing devices may adopt patterns of behaviour similar to humans [48, 55].

Shell used the exemplar of turn-taking to develop the principles of Attentive User Interfaces (AUIs) described here [48]. An AUI is an interface that measures attention and uses it to facilitate the turn taking process between the user and the device. This is accomplished by using multiple communication channels to interrupt or notify peripherally, as dictated by both the context of the user and by the content of the message to be delivered. In addition, an AUI can share its information and goals with other AUIs, and with remote people connected through technology. It manages the user's attentive resources by accentuating and underplaying information according to the user's attentive state. By passively measuring the user's attention, for example by recording gaze, the system can incur minimal additional mental load while providing additional services to the user.

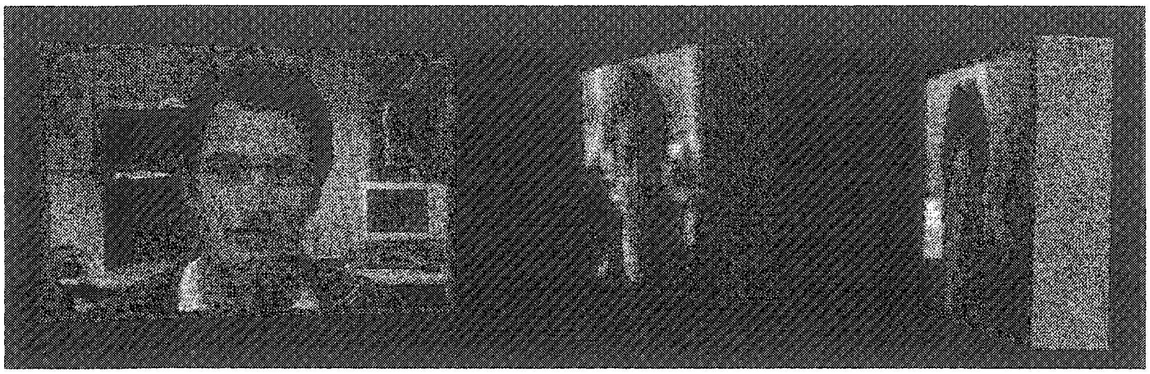

Figure 3. The GAZE-2 System. The two participants on the right are looking at the leftmost participant who is looking at the user. [48]

The GAZE-2 system developed by Vertegaal [58] communicates attention during remote interactions. Developed as an enhanced videoconferencing system, the GAZE-2 system texture-maps video images of participants onto separate faces of parallelepipeds (box-like objects) in a three-dimensional space. Each object on a user's screen represents one of the remote participants. An eye tracker is then used to determine which remote person the users are looking at. The objects that represent the individual participants in the videoconference are rotated by the computer to face each other in order to mimic the way users might look at each other (Figure 3). In this way, eye contact is conveyed to remote users, thus restoring this important turn-taking cue to video-conferencing.

Eyepliances [49] are household objects augmented with eye-contact sensors to respond to voice commands. These objects are networked together to communicate information about the current item being attended to by the user through the EyeReason server. Since users tend to look at objects before they issue spoken commands [33], the eye-contact sensor allows the augmented device to react to natural utterances while placing almost no additional mental load on the user. The eye-contact sensor allows the device to be aware that the user is addressing it. This simplifies the speech recognition problem by reducing the vocabulary that the computer must decipher to include only relevant words, thereby replacing the need for complex voice command syntax. Furthermore, the networked eyepliance's monitoring of the user's gaze can be used to determine the location and availability of the user for notifications by other devices.[2]

2.4 USABILITY

The consideration and pursuit of usability is central to human-computer interface research. The following section describes one definition of usability. This definition is used to illustrate the importance of the concept to the design of user interfaces.

2.4.1 The Definition of *Usability*

The definition and terms presented here are based on Shackel's definitions [47]. This source was chosen for its conciseness. Usability is defined in terms of four main principles: *learnability*, *ease of use*, *flexibility*, and *satisfaction*.

Learnability concerns itself with the ease with which a user can adopt a new system and attain a satisfactory level of performance. A system is easy to learn when it is *synthesizable*, that is, when it allows the user to build an accurate mental model. In addition, learnability is composed of *predictability*, *familiarity*, *generalizability* and *consistency*. Definitions for predictability, generalizability and consistency are straightforward: the interface should exhibit predictable behaviour, knowledge from one aspect of the interface should generalize to other similar aspects, and the interface should

[2] Horvitz designed Priorities [23], an example of an AUI that does not use gaze. The Priorities system rates received e-mails based on the average response time and frequency with which the user responds to the sender. High priority e-mails are then forwarded to the user's pager. The system manages the user's attention by forwarding only high priority e-mails, and works independently of the user's location by making use of a mobile pager.

be consistent in its reactions to the user. *Familiarity* refers to the presentation of functionality through the use of metaphors that are recognizable to the user.

Ease of use pertains to the user's ability to use the system to attain his or her desired goal. Important components of ease of use are *observability* and *responsiveness*. These criteria refer to the ability of a design to allow users to see the results of their actions and get immediate feedback. Also important are *recoverability*, as contained, for example, in an undo function, and *task conformance*: a criterion requiring that the user be able to accomplish the task he or she actually wants to accomplish.

Flexibility measures how well the system can adapt to users and their circumstances. Two contributors to flexibility are *substitutability* and *customizability*. *Substitutability* refers to the capability of the system to deal equivalently with input of the same type, that is, inputs that the users perceive ought to be interchangeable when they are of the same fundamental type. The *customizability* criterion asks that the user be able to alter aspects of the system to reflect individual preferences.

Finally, *satisfaction* criteria ensure that the user's experience with the system is pleasant. This criterion can be broken down further to reflect *emotional*, *aesthetic* and *environmental* attributes of the system: this last refers to the fit between the system as a tool and the task it is meant to assist.

Notable divergences from the definition of usability given above include the addition of safety criteria [42] and differentiation between short term learnability and retention over time [46]. The most significant departure from the definition is probably the removal of satisfaction from the definition of usability. In this case satisfaction is formulated as a distinct goal on par with usability [42]. The separation encourages a focus on satisfaction and on what users feel when working with a system.

The differences in emphasis in the alternative definitions are relevant when evaluating alternative approaches to design problems, as they suggest specific rules and measurements to take into consideration. These are then traded off when designing a system. As the aim of the research reported in this thesis is to develop the user's ability to engage with computers, it is important to understand the general goals of usability. The role that Auramirror can play in the design process of certain types of user interfaces will be discussed later. As a tool to aid in the design process, Auramirror is intended to place emphasis on the learnability and ease of use criteria. The criteria of familiarity and synthesizability are also invoked in the design of Auramirror.

2.4.2 The Importance of Usability

The importance of designing systems with improved usability follows directly from the early goals expressed by Bush and Licklider. The problem of achieving cooperation between humans and computers is inextricably linked to the problem of creating amenable and functional user interfaces. Indeed, symbiosis is not achieved when the individual entities struggle against each other to accomplish tasks. The importance of user interfaces to technology that will help humans is so fundamental that we can trace much of the history of computing through the development of user interfaces [36]. This is in part due to the relationship between the increasing functionality and complexity of technology as expressed by rapid hardware advances and the relatively constant capacity of humans to deal with complexity [7]. Improving usability can reduce the apparent complexity of a system so it does not surpass Buxton's *threshold of frustration* [7].

Frustration increases error and may cause a "general turning-away from technology" [37].

Finally, usability may affect the commercial success of a product by helping to develop more usable user interfaces. Future technology may enjoy greater commercial success more easily if Auramirror successfully plays a role in making the technology more usable. Donald Norman [37] argues that new technological developments might be commercially successful simply because they are sufficiently innovative, or provide functionality not otherwise available. In the case of mature technologies, where the functionality of devices and programs is well known and available from a number of competing companies, the success of a product depends on a very different set of criteria. Norman proposes that a shift to emphasis on ease of use, dependability and attractive appearance is required for the product to succeed in the marketplace.

2.5 PROBLEM DEFINITION

The main questions addressed by this research concentrate on the methods computers can use to recognize, measure, and then visualize the attention people pay to each other during face-to-face interactions. This problem statement refers to two general problems: the problem of asymmetry in human-computer communication, and the problem of information overload. Both these problems might be addressed by measuring and visualizing attention; each is described below.

2.5.1 Input/Output Asymmetry

The first problem is how to extend the input capacity for computers to improve the effectiveness of user interfaces, without inordinately increasing the mental load of the user. The current idiom of computer use entails an imbalance between the input and output capabilities of the interface: computers have extensive mechanisms for data output and restricted capacity for user input. Desktop computer users typically communicate their desires to contemporary processors through the constricted means of keyboard and mouse; the limited input bandwidth of these devices restricts the eloquence of the ensuing interaction. This, in turn, reduces the practical function of computers for many users who are unwilling or unable to communicate their problems adequately and, in this way, affects usability in terms of *ease of use* and *satisfaction*.

However, as promoted by UC and described by Vertegaal, the current idiom of computer use is changing [55]. Computers are leaving the desktop and moving into the environment. This presents an opportunity for technology to improve the input/output asymmetry by obtaining and exploiting larger quantities of input based on context-awareness. To this end, increasingly powerful processors can be linked to increasingly inexpensive sensors. Schmidt gives an overview of common sensors and how to use them to provide computers with general data about their context [45]. By using contextual information, it is possible for a system to obtain more input data with no additional work done by the user. The challenges for interface design then become what to measure and how to use the data.

One possibility is to measure the focus of the user's attention so as to create AUIs. Attention data can be used to direct interruptions to peripheral channels to avoid intruding on the user's current task, to transmit the attentive state of the user so other people can decide whether to interrupt or not and to what degree, and to distribute processing power to pertinent tasks.

2.5.2 Information- and Interaction-Overload

The second problem concerns how to improve the computer's communication to users by managing their attention more effectively through the use of peripheral channels and alternative methods of expression. With the proliferation of computers in the environment, human-computer interaction has fundamentally changed from a one-to-one to a one-to-many relationship [55]. As the numbers of processors around us grow, and the applications of computing solutions broaden, people increasingly find themselves at the limit of their abilities to manage the myriad devices in their immediate control. Paralleling the preponderance of computers is the increasing ubiquity of connections. It is common for a person to be simultaneously available via e-mail, telephone, mobile phone and instant messaging programs, while their computer and PDA reminds them of appointments and overdue tasks. In addition to requiring a greater timeshare of their awareness, the proliferation of computing devices and connections also affects social interactions. Already the interruption of a scintillating conversation by a mobile phone is a common occurrence and a common irritant. Today's users are bombarded with interruptions, regardless of context or conflict, which require immediate attention. The resulting competition between devices for the user's limited attention often results in information and interaction overload. A possible solution may be to introduce an awareness of attention to devices, so that a computer system might make use of cues that humans already use with each other in group settings. Attention is a relatively abstract concept. The challenge for this thesis is to find a way to make attention tangible in order for designers using attention, and scientists studying attention, to be able to visualize attention and directly assess this insubstantial quantity.

3
AURAMIRROR

Auramirror is a system that explores the means by which computers can estimate attention from contextual data and then visualize the attention people pay to each other. This chapter describes the aims of the system, the manner in which it depicts attention and its interaction with its (mirrored) audience. The merits of the system both in terms of functionality and an artistic pursuit are suggested. The chapter briefly discusses the inter-connections among attention, artistic representations of human-machine interaction, and user interfaces. Finally, the interrelationship between Auramirror and its participants is discussed, along with the issues of surveillance in the design of such technology. Implementation details are provided in the next chapter.

3.1 AURAMIRROR AND ATTENTION

Auramirror is an artistic installation that presents each participant's individual attention and the audience's shared attention through a computer enhanced video-mirror. The Auramirror images are a dynamic and artistic representation of the attention negotiations within the small groups of users that are captured by the camera. Auramirror's visualized attention provides feedback to a viewer to show how attention focus is shifting among the members of the audience. Auramirror creates an *attentive space*: a computer enhanced physical area, specifically a *perceptive space* that specifically attempts to measure attention and attempts to measure it passively as *context-aware* data.

Specifically, in this implementation, visible aspects of attention, such as looking behaviour, head orientation and presence, are measured for each member of the audience that is captured by Auramirror's camera in order to assess and then display, on a large wall-mounted monitor, the estimated attention audience members are exhibiting towards each other. Auramirror depicts a person's presence and *potential for attending* as an *attention aura* engulfing their head. The direction of an audience member's attention is depicted by a pseudopod-like extension to the attention aura. Auramirror augments the video image of the audience with the attention auras and projects the composite to the people in the room on the monitor's screen. The audience can then see themselves as though they were standing in front of a mirror with their attention focus being displayed for all to see.

Auramirror's visualization of attention allows its audience to assess whether or not the system is accurately reflecting the orientation of their attention and thereby offers a tool for studying attention as well as methods for identifying and understanding human interactions. Auramirror's measurements can be used to inform the design of user interfaces that need to sense a user's focus of attention and, in this capacity, work as a tool for fast prototyping. In addition, Auramirror can promote discussion about attention among its audience. It can open the eyes of the audience to the roles that computers already play in our attention spaces, and suggest the roles that they should play. The audience can provide feedback to help us advance our algorithms and methods for recognizing and monitoring shifts in our attention.

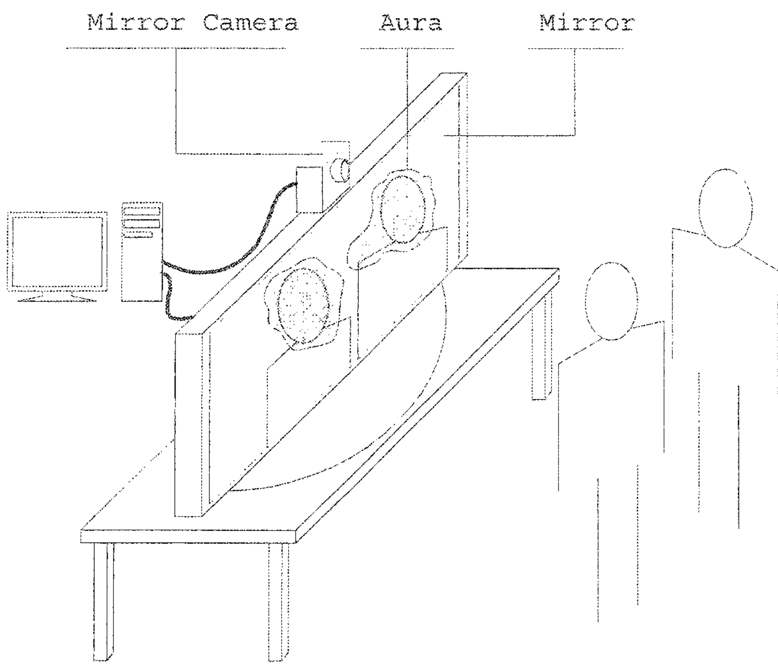

Figure 4. Configuration of single camera, two-person Auramirror implementation.

The Auramirror system uses a video camera to monitor an area in which an audience can enter (Figure 4). The system measures and tracks the estimated gaze of the members of the audience by using computer vision methods on the video image. In real-time, the system presents its deductions regarding the attention of the participants by superimposing blue blobs, the *attention auras*, over their heads in the video image, and then presenting the mirrored image on a monitor that can be seen by the audience (Figure 5a). The auras undulate around each person's head; this animation represents the dynamic nature of the visualization as a live reflection. The auras grow in the direction of their owner's attention as it is identified and estimated by the Auramirror system. Two auras will grow towards each other when the system determines the two people are focusing on each other. If the pair pay sustained visual attention to each other the auras will eventually merge to form an *attention tunnel* (Figure 5b). When their attentive bond is

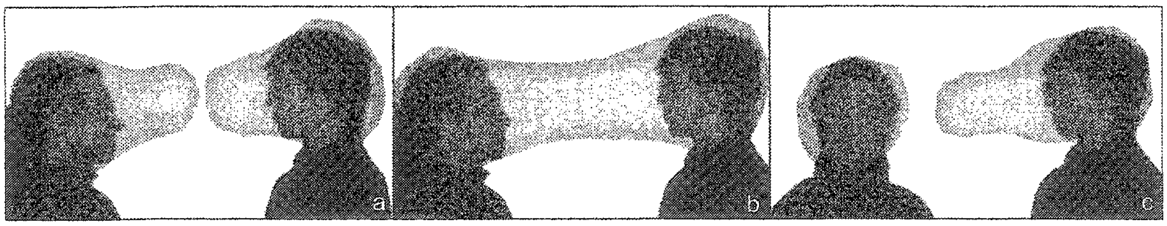

Figure 5. Stages of communication in Auramirror. Here we see the blobs unaffected (a), merging (b), indicating an attentive tunnel (c), and breaking (d)

interrupted, the tunnel snaps and the auras bounce back to their original rest positions around the respective heads (Figure 5c).

The following scenario describes a typical scene in front of Auramirror. Imagine that you walk into a room to face a mirror image, in a computer monitor, of yourself and of the people around you. You appear with a translucent, blue aura shimmering around your head. As you turn to your friend, she sees your head turn and your aura grow towards her. When she faces you to reply, everyone in the room can see the two auras merge to form the attention tunnel. Perceiving a movement in your peripheral vision, you turn your head and the people looking at the screen see the tunnel break and the attention auras spring back to their original shapes.

It is important to note that the position of the camera and mirror are chosen to emphasize the audience's awareness of how attention might be measured by a computer system and to promote exploration of what this implies. Depending on the function of the mirror other configurations may be used and additional monitors may be added. In other configurations the main mirror may or may not be visible to the audience.

Data visualization is realized in Auramirror by displaying the quantified values that it records on the location and orientation of faces. An attempt was also made to convey the *affordances* of attention to provide richer meaning to the data. The affordances of an object are defined by Gibson [19] and extended to design as *perceived* *affordances*, by Norman [38]. Affordances are the potential actions arising from the relationship between an observer and an object or environment. Perceived affordances are the potential actions that an observer is made aware of through the perception of the object or environment. In the context of data visualization, the perceived affordances are the apparent potential configurations and potential manipulations of the data expressed through its visualization. In communication, visual attention can be imagined as something that is projected or offered by one person and shared with another in measured amounts corresponding to the level of desired intimacy [3]. The visualization of Auramirror attempts to communicate these characteristics of attention that lie beyond the data, beyond the single points of focus. The auras show how people are enveloped by their attention space. The attentive tunnels show how people share their attention when they engage in conversation, and how they can include and exclude others with nonverbal cues that direct their attention and manage interactions. The attentive tunnels should evoke the sense of intimacy created by shared attention. By communicating these properties a deeper and more complex representation of the nature of attention is recognized. As part of a larger process, this broader understanding can be useful to people trying to harness the design space of attention. In terms of the technical definition of usability, Auramirror aims to be fast to learn and easy to understand.

3.2 THE USES OF AURAMIRROR

Auramirror can be separated into two components: a method for measuring the attention links of an audience, and a system for displaying this information. These two parts can be considered independently and the modularity allows us to modify the methods used to gather attention data without changing their visual manifestation, as generated by the system. The primary system described here uses estimation of gaze through head orientation as the method for measuring attention, and the display system shows the attention auras described above in a mirror image on a large wall-mounted monitor. Different algorithms for measuring attention can be used interchangeably with different display methods and configurations to find the best combination for specific situations. For example, different attention-measuring algorithms can be experimentally compared to indicate possible improvements.

Auramirror can be used without its display for attention recognition, and its monitoring capability can provide data to people studying group dynamics and the manner in which attention in negotiated. Understanding the group dynamics of attention can contribute to greater awareness by ubiquitous devices of their own audience. For example, such information can guide a perceptive environment capable of making decisions based on attention data. These decisions can help guide a system of many ubiquitous devices to act unobtrusively among a group of humans. The attention data can also be sent to a learning-capable system such as an intelligent agent. It can be used to help a system learn about the way attention is managed in groups of people and thereby improve the way the system interacts with people. Auramirror can be used to inform such a trained system while it is actively participating in group interactions. For example, FRED [57] is a software agent that uses gaze to help communicate with users. A greater overall understanding of group interaction could enhance FRED's ability to naturally partake in human interaction.

Auramirror exhibits information that can be used to evaluate algorithms for recognizing and tracking user attention. The ability to judge the accuracy of a method for measuring attention is important because of the difficult and subjective nature of attention. For example, attention may not be related to a person's gaze or eye movements when they are in deep reflection or contemplation of matters that have no immediate or spatial context, a phenomenon referred to as *orientation of thought looking* [26]. We can learn about the visual manifestations of attention from the feedback provided by people being monitored by Auramirror concerning what they were doing at the time the data were gathered and what they were actually paying attention to. In an informal setting, the real-time visualization capabilities of Auramirror can give quick results that allow fast iteration times and early evaluations to designers of attentive systems, thus helping improve fast prototyping of systems using attention. Auramirror does not require its mirror to measure attention. The mirror image can be recorded for later assessment without displaying it to the people in the attentive space, or the mirror can be positioned to be visible only to analysts outside the attentive space and not the participants. In this way, more traditional experimental methods can be used for appraising Auramirror's attention-measuring accuracy or studying other aspects of human communication.

Finally, Auramirror could facilitate collaborative work or serendipitous interactions between separated communities. The system could be modified to connect two separate locations through a live video connection in a public space, similar to the VideoWindow system [14], but with the additional power to render the attention of users through the mirror. With two networked installations at two locations, and with a modified mirror that displays the other remote location instead of a reflection, Auramirror can communicate attention from participants in one space to people in the other. This may help partially overcome the limited success reported by the VideoWindow system in promoting informal communication.

In summary, by reliably visualizing attention, Auramirror can help researchers studying the nature of attention and the criteria for measuring it. The methods used to validate Auramirror as a tool will offer a solid foundation for developing protocols for studying human attention through Auramirror.

3.3 THE ROLE OF ART IN HUMAN-COMPUTER INTERACTION

Art and design have always been closely related. Many designers in the field of HCI draw on art to help their products communicate, and to make them aesthetic and pleasing to people using them. Artistic techniques also play a role in the design process, and art has been the original inspiration for many design solutions. Works of art promote interaction with an audience by provoking discussion. As an art piece pertaining to attention, Auramirror quietly reflects its view of the attention people pay each other in a given setting, while also being an object capable of capturing the audience's attention. Auramirror also uses the role of artwork to provoke consideration of design issues regarding technology, attention, and surveillance. Auramirror was inspired by artistic expression, in addition to and enhancing its functional attributes. There are many ways by which art has influenced and inspired designs that improve communication with and through computers, and this section presents a few of the examples of the connection between art, attention and computers.

Garabet [17] posits performance art as a way designers can explore the effects of design decisions on the perception of the uses of devices. They use performance art to present wearable computers

> in a deliberately unusual manner where it is left up to the people interacting with the device wearer to imagine the intent of the device. By doing so, the focus is on reactions and comments of the public when they interact with the device and its wearer. [17]

Garabet and her colleagues staged three performances involving wearable computers with video camera and projector attached as peripherals. Each performance incorporated different design decisions. The system was worn in public where the live video was used to engage viewers who could see their own images being projected by the system. The performances solicited immediate reactions without the need for initial comments or encouragement by the wearers. The performances served to provoke reflection and debate about the device presented and its design.

Garabet intended to elicit reactions to the devices used in performances and the manifestation of design decisions. Specifically, the performances examine the values underlying the technology and the attitudes towards the devices. By using live video and loaded captions such as "Cameras reduce crime" the performances are intended to raise the issues surrounding surveillance. In response to a negative image people have of surveillance conducted by governments and corporations, Garabet concludes that wearable computing offers a potential for the individual to be the watcher and thereby become empowered by regaining some control over technology's potential uses for surveillance.

In another example, Dunne uses aesthetic considerations to instigate an awareness of technology's role in our lives [12]. Their Pillow is an LCD screen

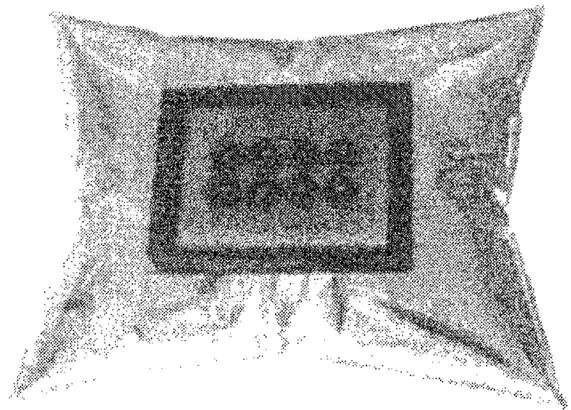

Figure 6. Dunne and Gaver's Pillow. [12]

embedded inside an inflatable pillow. The LCD screen changes its appearance to reflect ambient electromagnetic fields (Figure 6). The Pillow is intended to "raise issues about our existence in an omnipresent sea of electronic information" [12]. The display intends to raise our awareness of technology in our environment in a way that is made gentle by the pillow's beauty and subtlety.

The Projected Realities project [18] draws on ideas from art to influence the design process. The purpose of this project was to give a greater sense of presence to the elderly members in a particular community. First, cultural probes inspired by Surreal and Dada schools of art were sent to 10 elderly community members to begin the design process of the project. This approach used "tactics of ambiguity, absurdity, or opacity . . . to strip away habitual interpretations and open new possibilities." [18] Subsequently, conceptual art and contemporary works helped guide the design decisions of the group toward a final product: a way of automatically obtaining images and phrases from elderly members of the community and projecting these into nearby outdoor environments to convey the emotions and concerns of the elderly to the community at large. The designs that were developed by the group demonstrated "how technologies used on the periphery of attention can allow new forms of sociality to emerge." [18]

Artists often use techniques for directing the viewer's attention. Visual artists use lines, shapes, texture, colour and lighting to guide the viewer through their works (Figure 7). In some cases, characters in paintings employ non-verbal cues such as eye gaze to direct the audience. Glenstrup and Engell-Nielsen [20] described an experiment by Hansen and Støvring [22]: as the artist of several works, Støvring articulated the looking pattern of viewers that he intended in these pieces. Eye tracker recordings showed that subjects' actual eye-gaze patterns and movements were very close to what Støvring wanted and predicted.

The ability of the artist to manipulate attention is also supported by Wooding, who produced maps of looking behaviour for a number of paintings at the National Gallery in London [64]. These *fixation maps* show the locations of aggregate fixations over several volunteers for each painting; areas that receive a larger number of fixations receive a higher score. The resulting aggregate can be interpreted as the average looking behaviour of all the individuals recorded during the time span that the measurements were being recorded. The fixation maps were then applied to the paintings by covering each image with a transparent layer whose relative darkness is adjusted in different parts of the painting according to the accumulated number of fixations. Portions of the painting that few people looked at appear dark while areas that receive much visual attention appear normal. Figure 8 shows the results for Paolo Veronese's *Christ addressing a kneeling woman*. Notice how the looking behaviour of the characters in the painting seems to induce a particular looking behaviour in the audience.

Figure 7. Detail from Albrecht Altdorfer's *The Battle of Alexander at Issus* demonstrating how gaze can be manipulated by the artist. [65]

Figure 8. Paolo Veronese's *Christ addressing a kneeling woman* with brightness modulated according to the fixation map. [64]

Daniel Rozin [5] developed the Wooden Mirror as an art piece by attaching 830 pieces of wood to servos that could be tilted 30° up and down. Overhead spotlights illuminated the wood pieces, which individually reflected the light to produce different levels of brightness from different angles of inclination. A computer controlled the servos to adjust the tilt of the individual wood pieces in response to a processed video image from a camera hidden in the middle of the mirror. The Wooden Mirror demonstrated how computer systems can be used to develop interactive artworks that solicit and reflect the audience's participation. Wooden Mirror also demonstrated how the characteristics of a material can be extended by technology. In this case, wood, a material that is usually considered matte, is endowed with an ability to reflect images.

3.4 AURAMIRROR'S INVOLVEMENT IN THE *ATTENTIVE SPACE*

Using the metaphor of group interaction, Auramirror can play the role of a passive listener that a person is aware of at his or her attentional periphery. As a listener, Auramirror does not actively participate in conversations, but, as an observer, Auramirror's awareness of every audience member in the room is explicit, conveyed back to the audience by their images in the monitor. The audience is included in Auramirror's attentive space. Auramirror can act at the fringe of its audience's attention and can communicate attention information while staying in the background. By remaining at the periphery of the audience's awareness, Auramirror allows them to engage in regular interactions. At the same time, the negotiation of attention and patterns of turn-taking among participants can be observed, at the viewer's discretion, in

Auramirror's reflection. Like a calm technology the user can easily shift Auramirror between the foreground and background of his or her attention.

Although Auramirror does not overtly pursue the attention of the audience, it may affect it. The movements on the display may attract the attention of nearby participants in the view of the cameras; the contact shared between two participants might be interrupted if Auramirror gains an audience member's attention in this way. The largest and most exciting movements in the mirror are the formation of attention tunnels, and so the formation of the tunnel may draw a participant's attention to the monitors. It is when the people in the attentive space are just forming their most direct interaction that Auramirror has the greatest potential to interrupt the communication. In this way, Auramirror is involved in the audience's interaction even more closely by imposing itself through quiet interruptions at precisely timed moments. To a viewer who is susceptible to interruptions, the mirror has the potential to instigate iterative looking behaviour, elucidating attention and interruption and, thereby, mimicking turn-taking. This also suggests a way that computers can interrupt and interfere with normal behaviour, subtly and quietly, and without imposing a full awareness of its presence.

Auramirror may also promote interaction. Participants in the attentive space may be encouraged to communicate with each other to explore a system that responds to their interactions. In addition to provoking reactions to the interruptions of technology, the active qualities of Auramirror will reinforce the perception of the system as an attentive entity – a presence capable of affecting and effecting attention – indirectly a member of its own audience, able to affect its own display.

The duality of the Auramirror's role as an active participant and passive wall hanging elicits the mixed modality of people communicating remotely through devices. As a tool, Auramirror hopes to improve the way people can communicate through and with modern technology. As a visual statement, Auramirror evokes questions of what the difference is between attending to someone directly and attending to his or her image in the mirror. Technological advances allow us to connect with each other and with machines in ways not previously possible. Auramirror allows us to visualize this and encourages us to consider the implications.

Auramirror uses the theme of attention to help us examine the relationship between humans and computers. Auramirror can pay attention to the viewer by reflecting him or her in its mirror and by calculating the person's focus of attention. When the viewer leaves the Auramirror system the cessation of attending is mutual: Auramirror ceases to present the person in the mirror and the viewer stops attending to the mirror. Auramirror demonstrates the abilities of technology to function productively in our periphery. It also brings up issues surrounding the incursion of technology into our day-to-day interactions. Auramirror is designed to provoke an awareness of what a computer might measure. The data gathered by Auramirror are displayed publicly, i.e. Auramirror makes the entire audience more aware of the foci of attention of its participants. This information may be considered to be personal and private. (A husband, his wife and his secret or potential mistress should not visit attentive spaces at the same time). Some people in front of the mirror might become self-conscious of publicly displaying the object of their attention and their interactions may be affected by the unflagging, hyper-awareness of Auramirror. These possibilities suggest the potential power of conveying attention and emphasize the need for guidelines on how attention and other private information should be managed by current technology.

These issues provoked the deployment of the Auramirror prototype with naïve participants. Data were collected from surveys of users and results of the analysis are presented in the Chapter 5.

4
IMPLEMENTATION

This chapter describes how the exhibit at the Ontario Science Centre was implemented. The Auramirror system is described in terms of the individual components that comprise it. The environment setup describes the placement of the hardware components that are used for the system's input and output. The logic used by the Auramirror system is then outlined. A method for designing a system that would allow measurement of the attention foci in a larger group of people is also offered. The image processing algorithms used and the mechanics of the aura rendering are then presented. Finally the composition of these elements into the mirror is discussed.

4.1 HARDWARE SETUP

The Auramirror system implemented for the Ontario Science Centre used a single front view camera for both obtaining the video mirror image and for processing the activities of the audience (Figure 9). Processing the single video stream develops coarse data that can be used to estimate the attention between two people by monitoring their head orientation. A person's head orientation has been shown to correspond to their focus of attention 69 percent of the time [53]. For a full scene image, a participant's eyes will likely comprise too few pixels for effective measurement of gaze. Higher video capture resolutions may be possible with special hardware, but the analysis of these large images

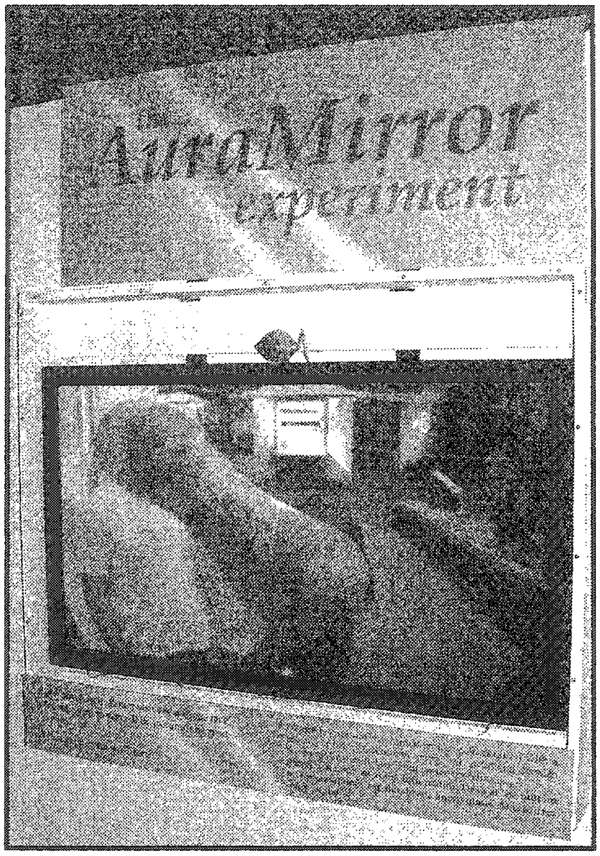

Figure 9. The Auramirror exhibit at the Ontario Science Centre. The instruction at the left reads, "Look at yourself. When you see a blob, face someone next to you. Talk to someone next to you. What happens to the blob?" The plasma screen image has been altered for demonstration purposes.

would impede real-time responsiveness. The advantage of the current method is in its simplicity and portability. Using a single video stream lightens the processing load and allows the entire system to run on a single computer while maintaining high frame rates.

The second method is illustrated in Figure 10 and potentially uses stereo imagery from one overhead stereo camera or from two calibrated overhead cameras, in addition to the frontal mirror-camera. The additional processing load of this configuration may require more than a single computer to achieve real-time performance. The first method has been implemented and the second is included in this discussion to show how Auramirror might be scaled to track the attention of more than two people. In both setups, the video image of the audience would ideally be untainted by parallax, the difference from the expected mirror image due to the position of the camera, and reflect as close to full scale as possible. To approach these criteria, a large plasma monitor was used and the camera was placed directly on top of the monitor. This position produces a less noticeable parallax effect than do side or bottom cameras [8]; it also allows a more flattering angle of the audience. For the exhibit at the Ontario Science Centre, the Auramirror software ran on a Pentium 4 3GHz processor. The video was captured at 320x240 pixel resolution with an IEEE 1394 camera. Video capturing was done using DirectX 8.1's DirectShow [35]. The mirror was rendered using OpenGL [40] on a 62-inch plasma screen, as illustrated in Figure 9.

4.2 THE MIRROR LOGIC

The first, single-camera method by which Auramirror tracks and estimates attention uses a computer vision classifier on the camera image to find faces in the scene. The image processing algorithms will be described in Section 4.3. Once a face is found, Auramirror spawns an aura at the face's location in the image and initializes mean shift colour tracking of the audience member's head. Using the face finder and the colour tracking location, Auramirror determines when someone is facing the mirror by detecting a frontal face. If neither participant is facing the mirror they are labelled as potentially facing each other. Counting these states over a number of frames will produce a rolling average that can smooth state switching to avoid noisy jumps due to inaccurate recognition. When two heads are estimated to be facing each other, the system identifies this as mutual attention. This triggers the attraction of each aura towards the other, which will be described in section 4.4. This animation terminates with the creation of an

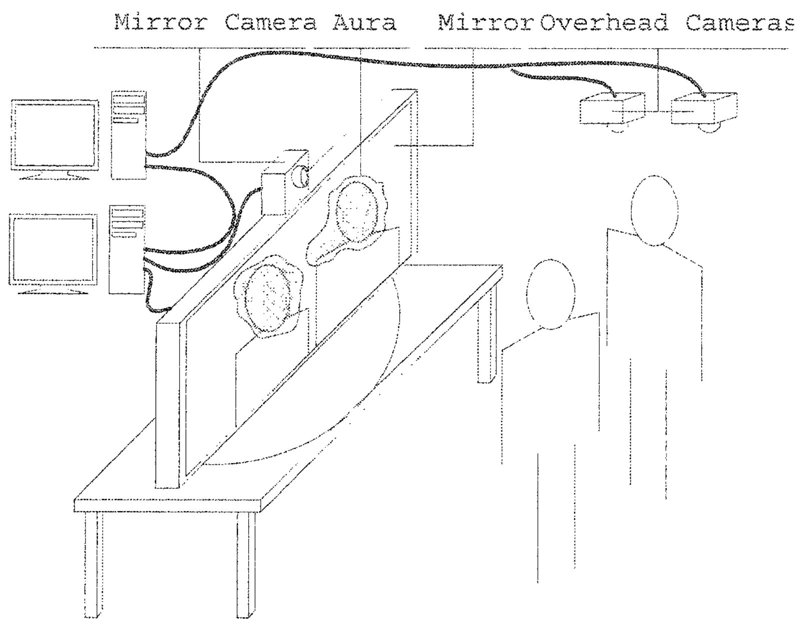

Figure 10. Multi-camera setup of Auramirror.

attentive tunnel, a single aura that replaces the individual auras and that covers both heads.

For the system to identify a new audience member it must first find a frontal face. In this implementation, people tend to approach the system while looking into the mirror, and so tend to be facing the camera. This was the case at the Ontario Science Centre to such a degree that a notice to encourage people to look at each other while in front of the system was added (Figure 9). Unless they did this, full interactions would not be identified by the system and the audience would not see the merging action of the auras. Once people are discovered they are primarily tracked by their face's colour distribution. If a face is found within a preset distance of the colour's reported position, the system concludes that the face belongs to the person being tracked. When this happens, the colour tracking position is adjusted and reinitialized. This improves the robustness of the ongoing tracking procedure and makes it difficult to manipulate an aura so that it tracks something other than the face.

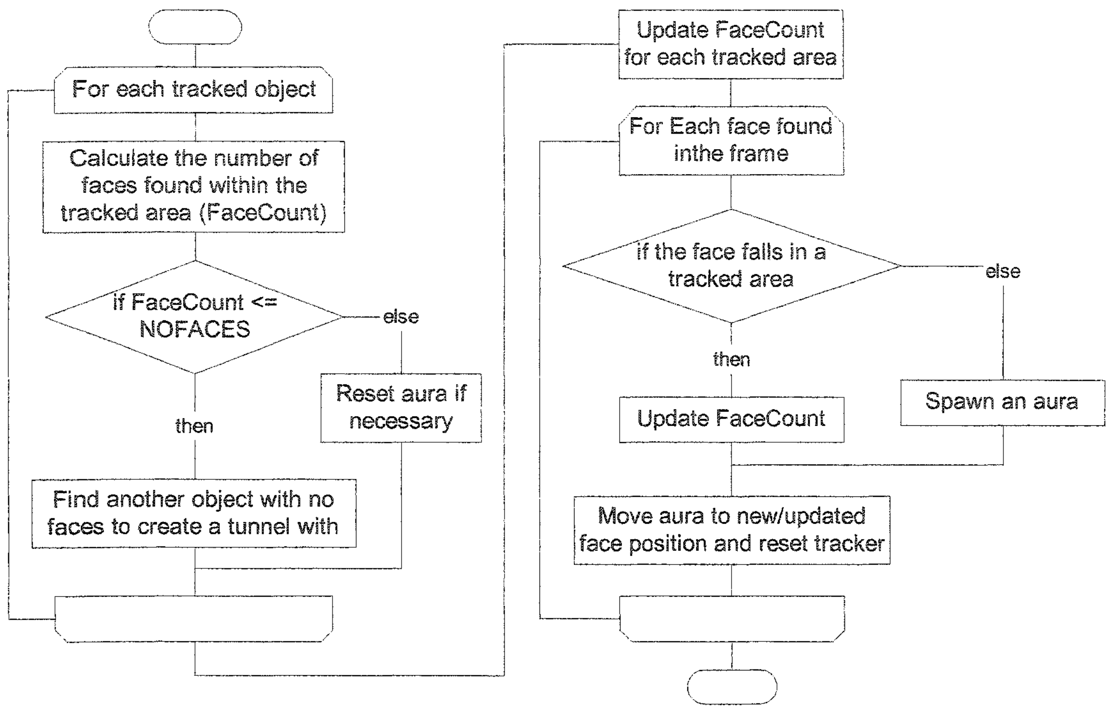

Figure 11. Flowchart for single camera prototype.

The system must also determine when to remove an aura that belonged to a person who has left the scene. There are two methods for doing this. The primary method is through the colour tracker. The colour tracker algorithm iterates, refining its position with each iteration, until either the position is determined to be stable, i.e. the change in distance between iterations is below a set threshold, or until it reaches a maximum number of iterations. Reaching the maximum number of iterations usually indicates that the colour to track has been suddenly removed, for example, if it disappeared off the edge of the view. The aura is dropped when this occurs for a set number of consecutive frames. This behaviour is exhibited usually when a participant leaves the scene by passing out of a side edge of the system's view. However, when a participant moves further away from the camera and then moves out of the scene, the tracker can sometimes stabilize on a background object and fail to report maximum iterations. For this scenario, the second method of detecting an abandoned aura is used. When the aura's cumulative movement over several frames is too small, the aura is removed. In either case, the aura will not be removed if a face was recently found within the tracked region. The logic for the single camera implementation is summarized in the flowchart shown in Figure 11.

The second possible method to track and estimate attention would use two overhead cameras that are calibrated for stereovision to capture images from a larger and more complex field. From the overhead view, an adaptive threshold of the depth image as well as a size threshold can be used to find the heads in the scene as connected components. This method may partially resolve some of the problems created by using the frontal camera when more than two people are in the attentive space when, for example, heads obscure each other. With heads found, the system can hypothesize that the head orientation is in the direction of the major axis of the ellipse that best fits the head's connected-component. Auramirror might use skin colour detection from the mirror scene camera to disambiguate the two possible directions of attention provided by the major axis vector. This vector can then be extended to test for possible intersections with other head connected components. A positive intersection indicates attention towards the associated person. In this scenario, the aura placement and animation will predominantly be directed by analysis of the overhead camera images. The mirror camera will primarily reflect the scene and overlay the auras.

Using the overhead cameras for range data may broaden the systems ability to track a larger group of people in a larger area. With this data, a variety of algorithms for tracking and identifying objects can be applied for the same benefit. Research of more complicated methods for object recognition such as the work done by Greenspan [21] can be applied to identifying heads and head orientations from this type of data.

4.3 IMAGE PROCESSING ALGORITHMS

The programme code for the Auramirror software was written in the C++ programming language. The DirectShow Software Development Kit (SDK) is a framework used for programmatically managing the capture and rendering of video and sound on a computer. Intel's Open Source Computer Vision (OpenCV) Library [39] is an open-source programming library for image processing. The Auramirror software employed functions and data types from this library. The four algorithms described in this section were built into three DirectShow filter classes, one for face detection, and one for multiple Kalman filtering using *mean shift* colour tracking, and one for background subtraction. The algorithms are executed in the transform methods of the filter classes, which are usually executed in a separate thread from the main application. Each filter begins by converting the frame buffer received by the filter to the OpenCV image format and indicating that the transform method is currently executing, to avoid race conditions on output parameters. DirectShow filters are further described in Section 4.5.

4.3.1 The Haar-Feature Face Detection Algorithm

The main face detection algorithm was implemented in an existing OpenCV function that is called from the transform method of a DirectShow filter. This section will give a brief overview of the algorithm. A complete description can be found in [32].

To initialize the face detection function, pre-calculated values that correspond to a trained classifier are loaded. OpenCV provides values for frontal face detection based on extensive training. The classifier consists of multiple stages. Each stage of the classifier is trained using the Discrete Adaboost algorithm [15]. This algorithm uses a technique known as boosting to find complex criteria for the classification of data based on

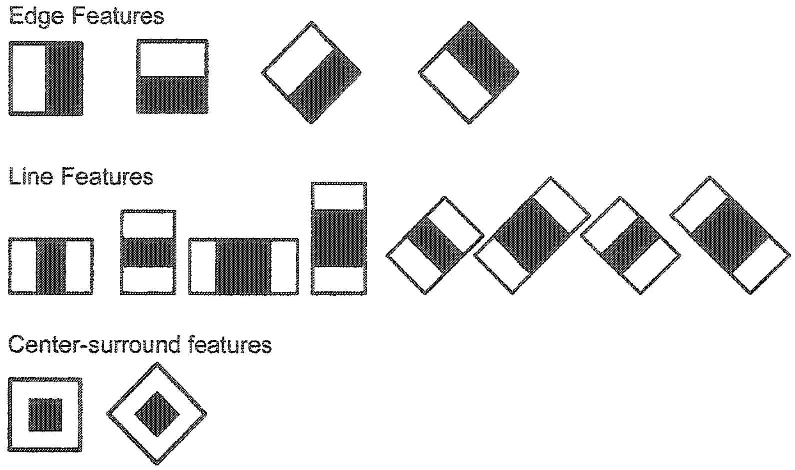

Figure 12. Extended set of Haar features used for building the classifier.

combinations of *weak classifiers*, a collection of very simple criteria for classification. In this case the algorithm uses, as its set of weak classifiers, wavelets from the extended Haar set (Figure 12), and the combinations are described by spatial layout and sizing of the multiple wavelets. The multiple stages of the classifier compose a degenerate decision tree: an input must pass all the stages successfully to be accepted, this is the *cascade*. Each stage is trained to detect all positive samples, to reject some percentage, X, of negative samples and is limited to allow some small percentage, Y, of false rejections. For a classifier with N stages, this results in X^N false alarms and a $(1-Y)^N$ hit rate. If the percentage of input samples that contains faces is given by P, then the theoretical error rate is

$$\text{Error rate} = P[1-(1-Y)^N] + (1-P)X^N. \qquad (1)$$

The features used as weak classifiers are shown in Figure 12; each one is a white rectangle containing some black area. These wavelets are meant to approximate fundamental graphical units such as spots, edges and lines. These wavelets can be scaled, translated and superimposed to reconstruct approximations of greyscale images. To evaluate an image, for example during training, the features are scaled and translated to cover different areas and the pixel values over the covered area are summed as well as the pixel values over the dark area of the feature. These two values are subtracted, and if the result is close to zero, as determined by a preset threshold, then the feature is considered a good descriptor of the area it covers. These computations can be done in constant time after a single-pass pre-computation of the image data. A combination of features with their associated positions and scaling are learned for each stage of the classifier. Subsequent stages are trained on those negative samples that pass undetected from earlier stages. As a result, the complexity of the classifier for each stage tends to increase with its depth, and so the longer, more convoluted classifiers are only evaluated for samples that have passed the earlier tests. For a more detailed description of the methods used to train the classifier see Lienhart and Viola [32,60].

To find an object in an image, windows of increasing size slide over the image. The classifier assesses the resulting sub-image in the window, stage by stage. A rejection from any stage rejects the sub-image. Because of the cascade, the classifier can quickly reject unlikely sub-images refused by early stages of the classifier.

4.3.2 The Multi-Kalman Tracking Algorithm

The basic discrete Kalman filter works by recursively adjusting its predicted state estimate, in this case the location of the face, with a state measurement, in this case the colour-tracked location returned by the *mean shift* algorithm. This section will give an overview of the Kalman filter based on the discussion in [63] and briefly describe the mean shift algorithm. For more detailed information on the Kalman filter see [27,63]. For more detail on the mean shift algorithm see [16,10].

The filter iterates for each frame processed by the system. Here, the final estimated state is denoted by $\hat{x} \in \Re^n$, the *a priori* estimate $\hat{x}^- \in \Re^n$ is the final estimate from the previous iteration. The measurement from the current iteration is given by $z \in \Re^m$ and the unknown, real state is $x \in \Re^n$. The final estimated state reported by the Kalman filter is given by a linear combination of the *a priori* estimated state and the residual: a measure of the difference between the predicted, $\mathrm{H}\hat{x}^-$, and measured state,

$$\hat{x} = \hat{x}^- + \mathrm{K}(z - \mathrm{H}\hat{x}^-). \tag{2}$$

Here the matrix $\mathrm{H} \in \Re^m \times \Re^n$ derives from its relation of the actual state to the measured state and noise, ν:

$$z = \mathrm{H}x + \nu. \tag{3}$$

In Auramirror the measurement is meant to exactly reflect the actual state, so H is simply the identity matrix. The matrix $\mathrm{K} \in \Re^n \times \Re^m$ is the blending factor, and is chosen to minimize the *a posteriori* error covariance given by

$$P = \mathrm{E}[(x - \hat{x})(x - \hat{x})^T]. \tag{4}$$

For the Auramirror software, the OpenCV Kalman filter implementation had to be modified in two ways. First, it was changed to allow for simultaneous tracking of multiple features. Second, it was altered to use the Mean Shift algorithm for its measurement phase instead of the default algorithm used in OpenCV. In addition, some logic was added to prevent the trackers from tracking the same item.

Although the tracking was not as robust at sticking to its target, especially in low light conditions, it was found that the Mean Shift algorithm was better at indicating when a person had left the scene than OpenCV's default algorithm. The Mean Shift algorithm starts with a colour histogram of the area to be tracked, in this case a face. The algorithm then iteratively finds the best match to the object colour histogram in the frame, while minimizing the distance from the original target center. The algorithm iterates until either the change in the found location is lower than some threshold, or until some maximum number of iterations is reached. If the maximum number of iterations is reached, this is reported. If this happens for several consecutive frames, the main program may determine that the object has left the scene.

Finally, since each tracker is meant to follow a single person, programme code had to be added to prevent the trackers following identical regions. The program iterates through all the currently active trackers and calculates the mean shift for each one at a time. Before the measurement phase for one tracker is started areas from other trackers are set to black if they overlap the current tracker's area. The Mean Shift algorithm is robust enough that if people temporarily occlude each other they will usually be picked up when they reappear – though possibly after exchanging trackers with their "occluder".

4.3.3 The Background Subtraction Algorithm

Currently a participant's presence is first registered only when their frontal face is recognized by the system. In the hope that the necessity of detecting a frontal face for initialization could be avoided a filter to subtract the background image was created. Significant areas of deviation from the background could result from new people entering the camera's field of vision, and so trigger the systems registration of a new participant.

However, this method would complicate the initialization process of finding specifically a head and tracking it. Also, this algorithm was computationally intensive, resulting in a significant lag in the system's performance. Finally, this method is subject to error due to camera movement, including vibrations, and lighting changes, though there are more complicated ways to account for these. For an example of how this approach might work see Wren's Pfinder system [67]. Since people tend to look at the mirror as they approach, and the face finding must be done regardless, there is a viable alternative to using background subtraction. Therefore, this algorithm is not used in the current implementation of Auramirror.

The algorithm that was developed for background subtraction begins with a several frame initialization, during which the scene must contain only the background. During the initialization period, the sum of the pixel values and the sum of the squares of the pixel values are computed. From these numbers and the number of frames used for initialization, the mean and standard deviation are derived for every pixel in the image. Thereafter, each frame compares its pixel values to the mean value and the corresponding pixel is set in an image mask if it differs by some factor of the standard deviation, typically three standard deviations. Either the comparison can be done with the red, green and blue (RGB) colour components of the input image, or the image can be converted to hue, saturation, and brightness (HSV) format and the comparison can be made on a subset of these channels. The conversion adds more computational complexity but in informal evaluations seems to slightly reduce the chance of detecting shadows of objects as changes in the scene background. This is because the hue and saturation change less because of a shadow in the image; a shadow primarily affects the brightness channel. The algorithm requires several simple calculations per pixel per frame. Although the time complexity is linear in the number of pixels, the significant number of operations and the large number of pixels make the delay due to this filter noticeable. Also, noise produced in the image for imperfect background subtraction drastically reduces the accuracy of results from the face finding algorithm.

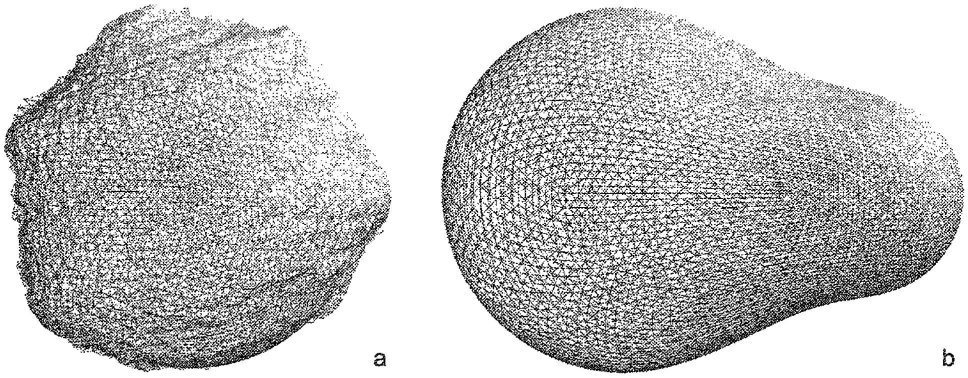

Figure 13. Perturbed (a) and Attracted (b) meshes.

4.4 AURA MECHANICS

The auras are the visualization of attention provided by Auramirror. They are modeled as elastic, spherical, triangular meshes. The aura functionality and rendering are compiled in the blob function library, authored specifically for this research. The mesh is created as an icosahedron-based geodesic sphere using class I subdivisions [28]. The frequency of subdivisions determines the number of faces and can be adjusted to achieve the desired balance between animation responsiveness and physical smoothness. In addition to simple translation and scaling the aura mesh can be manipulated in two ways: it can be *perturbed* or *attracted* as shown in Figure 13. The ensuing mesh animation is either a static translation of individual vertices, or a simplified and contrived simulation of energy dispersion across the mesh. Note that although the purpose of these methods is

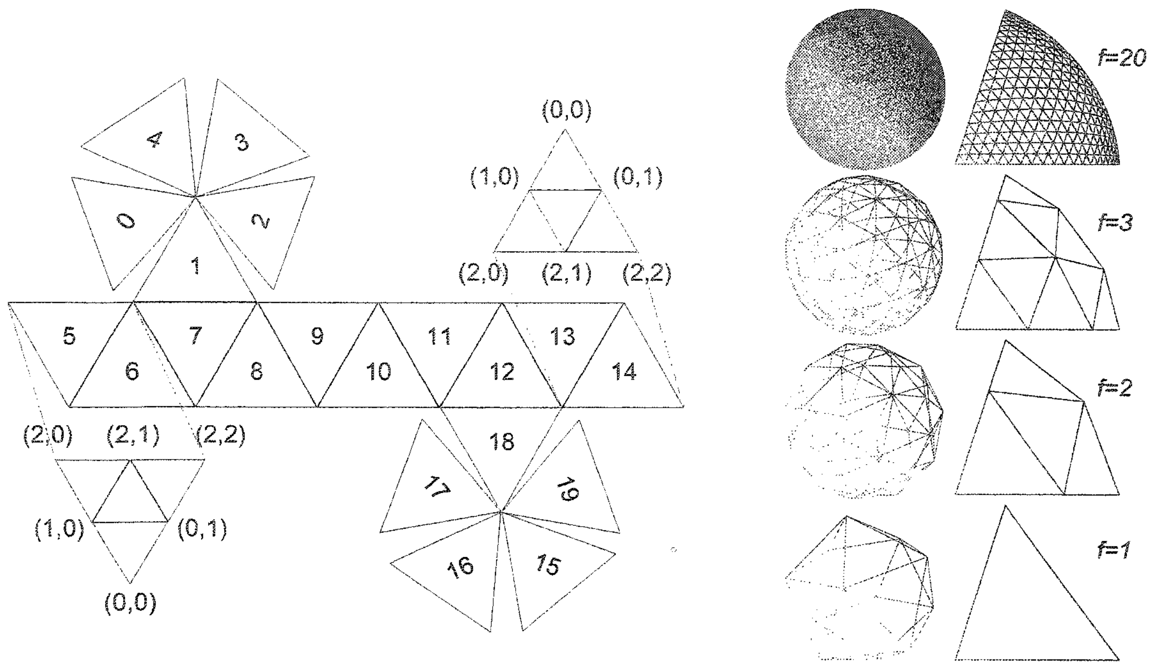

Figure 14. On the left indices for faces in the face-graph (black) and subdivision coordinates (blue insets) for upward and downward pointing faces with frequency 2. On the right the effect of increasing the frequency (*f*) can be seen for the icosahedron and for individual faces.

to convey the fluid nature of the auras, these are not attempts at an exact mimicry of nature. In addition to run-time efficiency considerations, the auras must also be able to be manipulated in ways that are unnatural to most fluids.

4.4.1 Aura Data Structure

To simplify computations on the mesh several data structures are maintained. In addition to Cartesian vertex coordinates, the vertex rest positions are stored in Cartesian and spherical coordinates as well as the normal vector, rest normal vector, and velocity. Coordinates are local, relative to the mesh centre. Pointers from vertices to edges and faces and vice versa are also maintained. Edges, in addition, have their rest length recorded; this is because edges of a geodesic are not all equal. These data structures are populated using a temporary structure called the face graph. The face graph uses three coordinates to point to uniquely numbered vertices: the face number, and two coordinates indicating the position of the vertex relative to the face (Figure 14). The structure is redundant, as faces share vertices. The faces of the original icosahedrons are numbered, winding down from the zenith (Figure 14). By carefully considering the order of assignment, the spatial coordinates can be assigned to each vertex as the face graph is populated. Afterwards, the lists of edges and faces are generated.

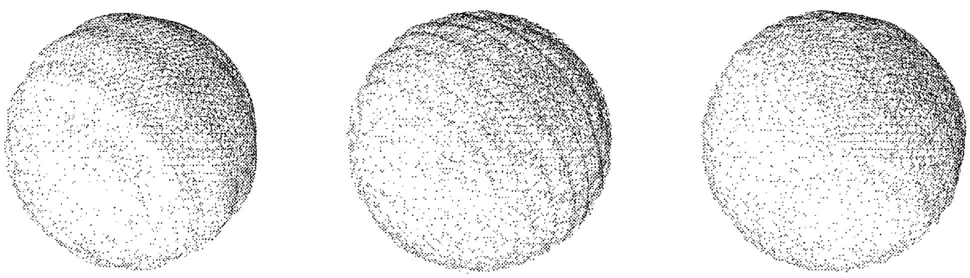

Figure 15. A sequence of frames showing the perturbation of an aura mesh. (Frequency = 20, $\alpha = 0.2$, $\beta = 9$)

4.4.2 Perturbation Algorithm

The purpose of perturbing the mesh is to make the display dynamic and to communicate the fluid nature of the auras. At regularly timed intervals, the vertex positions of the mesh are given velocity along their normal, based on distance from two randomly selected vertices. The exact relationship is,

$$V_i = V_i + \alpha \sin[\beta(\frac{\|P_i - P_a\| + \|P_i - P_b\|}{2\|P_a - P_b\|})]N_i \,. \tag{5}$$

Here V_i represents the velocity vector of the i^{th} vertex, P_i represents the position vector, N_i represents the unit normal vector of the vertex, a and b are the indices of the randomly chosen vertices, β varies the frequency of the perturbation wave, and α is a constant to vary the scale. α and β are given random values for each perturbation. The undulations caused by the added velocity contribute aesthetically to Auramirror. Also, the

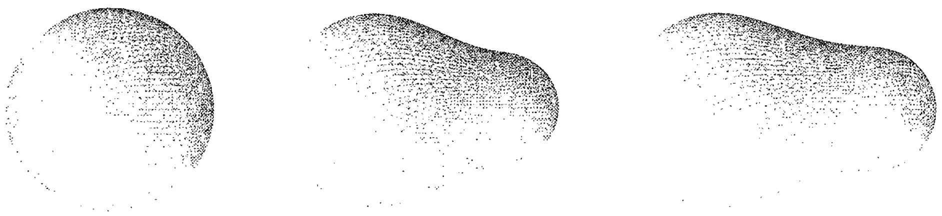

Figure 16. A sequence of frames showing the attraction animation of a single mesh. The droplet sphere grows as it moves to the right. (Frequency = 20)

fluid and dynamic nature of the auras visually engages the audience and encourages them to deduce the meaning of the exhibit.

4.4.3 Attraction Algorithm

When the system determines that two people are paying attention to each other, it instigates an attraction sequence that stretches the auras in the direction of each other until they are close enough to merge. In this early version of Auramirror, object collision is not computed. As opposed to perturbation, in which the aura always relaxes towards its original shape, the attraction method will alter the rest positions of the vertices. This allows the aura to take on and retain configurations that would otherwise be unstable. The method used to deform the mesh to indicate attraction is based on a separating droplet model. An attracting aura is composed of two spheres: the main aura sphere, centered at the original mesh's sphere centre, and a smaller droplet sphere some distance away. To animate the attraction, the smaller sphere moves away and its radius alters over some fixed number of frames (Figure 16).

New vertex positions are determined by computing an offset for each vertex of a sphere superscribing both the aura and droplet spheres. This method ensures a more even distribution of vertices across the attracted aura's surface, which in turn improves the aura's appearance. For each of these vertices on the supersphere, we find the vector to the corresponding position on both the aura and droplet spheres, points *A* and *B* in Figure 17.

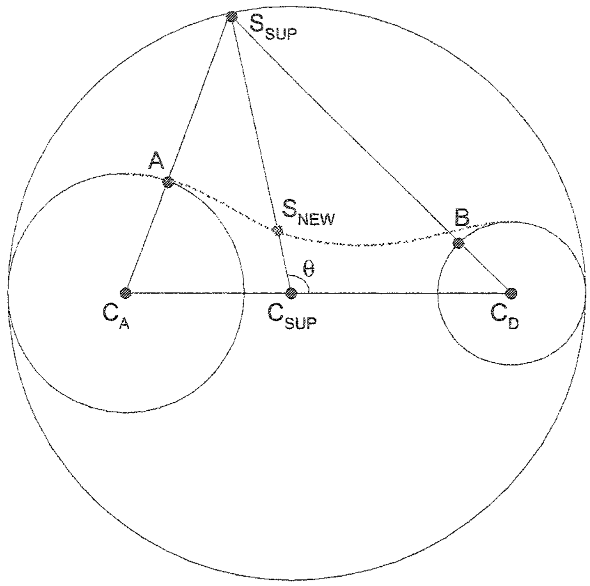

Figure 17. A cross section of the spheres involved in computing the attracted mesh. C_A is the aura centre, C_D is the droplet centre, S_{SUP} is the current iterated vertex on the superscribed sphere. *A* and *B* are points on the aura and droplet respectively. The vectors *SA* and *SB*, and the angle θ is measured. S_{NEW} is calculated.

With the calculated information we compute the new vertex position, S_{NEW}, by mixing the influence of both spheres according to a sigmoid function of the cosine of the angle θ in Figure 17.

$$S_{NEW} = S + (1 + \mathrm{K})\overrightarrow{SA} + \mathrm{K}\overrightarrow{SB}, \tag{6}$$

where,

$$\mathrm{K} = \frac{1}{1 + e^{-\alpha \cos\theta}} \; . \tag{7}$$

4.4.4 Mesh Dynamics

The mesh dynamics algorithms determine most of the animation of the auras. Vertices will always be drawn to their rest position, to such a degree that in some circumstances it appears unnatural, but this prevents people from having their auras "boil away" from too much energy, for example accumulated from excessive participant activity. Note that all constants denoted by α are constants set in the code header. In this section they each refer to an efficiency percentage. The vertex calculation for one frame begins by measuring the potential energy of each vertex, *PE*, which is simply taken to be the vector of the current vertex position to its rest position. If we define the scalar $\Delta PE_{i,j} = \alpha_V \|PE_j\| - \|PE_i\|$, where α_V is a constant roughly denoting efficiency due to viscosity then potential energy is gained from neighbouring vertices as follows,

$$PE_i' = PE_i - \sum_{j \in A(i)} \frac{\Delta PE_{i,j} N_i}{\|A(i)\|}. \qquad (8)$$

Here $A(i)$ is the set of vertices such that each element, j, shares an edge with vertex i and satisfies the condition $\Delta PE_{i,j} > 0$. Equation 8 gives a new potential energy, PE_i', along the vertex normal, N, based on the potential energy gained from neighbouring vertices modified by viscosity. The new vertex velocity, V', is calculated based on the change in potential energy, modified by a conversion rate α_{CV}, and the old potential energy modified by elasticity, α_{EL}, of the mesh and the previous velocity, V,

$$V' = V - \alpha_{CV}(PE' - PE) + \alpha_{EL} PE. \qquad (9)$$

Note that where potential energy is calculated for each frame, velocity persists from frame to frame. Offsetting the vertex rest position, R, by the new potential energy and the new velocity, determines the new vertex position, P,

$$P' = R - PE' + V'. \qquad (10)$$

Once the new vertex position is calculated a final value, $\hat{V}$, is figured and retained for use at the next iteration, the next frame. This computation will reduce the velocity's magnitude by potentially two factors. The first always-applicable factor, α_F, is due to general inefficiency, for example due to friction. The second factor, α_D, is due to the hypothetical resisting force from compression of the aura's volume. The second factor is only applicable if the vertex position is closer to the aura centre than its rest position, and so is inside the usual aura perimeter.

$$\hat{V}_i = \alpha_F f_{\alpha_D}(i) V_i' . \tag{11}$$

Here $f_{\alpha_D}(i) = \alpha_D$ if the vertex i is closer to the mesh center than its rest position, otherwise $f_{\alpha_D}(i) = 1$.

4.5 AURAMIRROR SOFTWARE

The components described above were brought together in a central program that was responsible for initialization, for managing user input such as keyboard hits, and for implementing the mirror logic. The principle challenge was getting the Component Object Model (COM) -based DirectShow code to work in conjunction with the OpenGL-based GLUT (OpenGL Utilities Toolkit) execution cycle.

Microsoft provides a SDK for DirectX that includes DirectShow. DirectShow is a COM-based framework for managing input, processing and playback of video and sound. At the centre of DirectShow are the filter and the filter graph classes. Each filter object is a component that accepts input and produces output by manipulating a data buffer in its transform method. Conceptually, a filter usually performs a single operation on a multimedia stream. Filters are connected together to form a filter graph, in the case of Auramirror the first filter will be one that reads video input from a camera via a graphics card driver; the last filter is usually a rendering filter that renders the data to a monitor or

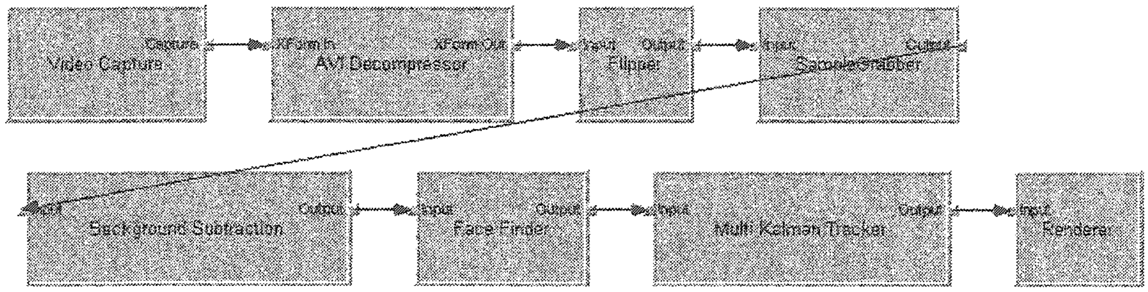

Figure 18. Filter graph for Auramirror.

a file, or in the case of the null renderer, simply drops the data. For the filter graph used by Auramirror see Figure 18.

Each of the filters created for Auramirror are in-place transform filters. This means that the data is worked on, in-place, in memory; DirectShow does not require you to copy the buffer to a new place in memory. The main function of an in-place transform filter is carried out in the classes Transform method. The filter graph runs as a streaming thread separate from the application thread, with each filter potentially running separate worker threads. In several filters created for Auramirror, data are provided to the main application thread. As a result, care must be taken to synchronize access to variables whose values are shared. All the filters created for Auramirror perform two operations in the transform method: they set a Boolean value in a critical section to indicate that the Transform method has been entered, and they put the frame buffer containing the current camera image that is passed into the method into an OpenCV image format. This is a simple matter of setting appropriate header values and pointing the data pointer to the buffer. Field values accessed by methods called by other threads are only set in the Transform method within critical sections.

OpenGL Utilities Toolkit (GLUT) is a set of functions designed to simplify the task of creating OpenGL rendering applications in a Microsoft Windows environment. The GLUT main function initializes the call-back functions that will respond to operating system events such as window drawing, keyboard hits, and window resizing, and then starts the execution cycle. The execution cycle calls pertinent events in an infinite loop, replacing the usual windows message pump typically used by the Microsoft Foundation Classes (MFC). Since the DirectShow filter graph is designed to run as a separate thread, care must be taken to ensure that the GLUT message loop does not interfere with the DirectShow thread execution and vice versa. In addition, the DirectShow code uses MFC, which, if poorly configured, can pre-empt the GLUT message loop. It is exceedingly easy to create an OpenGL window that never receives any data from an error-free, running filter graph!

To properly get OpenGL to work with MFC, COM, and DirectShow, the application thread for Auramirror is a console application. This prevents the MFC windows message pump from conflicting with the GLUT message pump by avoiding it altogether, and allows smooth spawning of the filter graph thread.

4.6 SUMMARY

The Auramirror software combines image-processing algorithms into DirectShow filters to process live video images. A central program receives data from the filter graph thread to inform the logic of aura management. This program uses the blob function library written for this project to render attention auras and attention tunnels into an OpenGL window. The program also receives the current unadulterated video image data from the filter graph. The image is texture-mapped onto a background object placed in the OpenGL space, causing the video to fill the background of the OpenGL window. Informal tests of the system indicated that it works as intended: frontal face finding seems to be a good indicator of when people are looking at the mirror, and the mean shift tracking follows people's heads. It seems that lighting is important for these methods to work well since they both rely on sufficient colour contrast. The results of a formal evaluation of the system are presented in Chapter 5.

5
USER REACTIONS

This chapter describes and presents the results of a survey of participants that was conducted to evaluate Auramirror. A prototype of Auramirror was deployed at the Ontario Science Centre in Toronto. Questionnaires were distributed at the exhibition during a one-month period to gather data on the participants' interpretations and reactions. A total of 121 completed questionnaires was obtained. This chapter presents the goals of the survey and the methods used in the evaluation; the results are subsequently presented, and the chapter concludes with a preliminary user assessment of the system.

5.1 THE METHOD

A questionnaire was designed to gain descriptive data on the reactions of the people who came to see and participate in the Ontario Science Centre demonstration of *The Auramirror Experiment*. The first two goals of the survey were to determine if the participants thought that Auramirror's attention-measuring algorithm was accurate in estimating the focus of attention, and also to find out if the algorithm was effective in identifying interaction between participants. The third goal was to assess the claim that Auramirror's method of visualization was effective at conveying its meaning as a representation of attention. To respond to this objective it is important to ensure that the interactions recognized by Auramirror are reported precisely. The ease with which the visualization of attention in the mirror is understood by participants is important in the evaluation of the system. Ideally, Auramirror is intended to display attention in a way that people can intuitively understand, and in a way that permits the perception of the potential for attention to be directed between participants. Ideally participants perceive the intended affordances of attention.

The survey was also used to gain feedback on participants' relationship to Auramirror. The survey endeavoured to measure if the system was encouraging interaction between people who did not previously know each other. The survey also attempted to measure whether or not people enjoyed using the system, and if they tried to manipulate the auras deliberately. The survey addressed the issues of surveillance to find out if people felt that they were being watched and made to feel uncomfortable.

Auramirror was installed in the Communications Hall, Level D, among other Ontario Science Centre exhibits. Instructions at the exhibit encouraged people to turn and face each other so that could see the auras merging and connecting to form tunnels (see Figure 9 from Chapter 3). When an Ontario Science Centre host was available, visitors were personally asked to fill out the questionnaire, otherwise instructions were left at the exhibit asking people to go to the information desk in the next room to get the one-page questionnaire. This biases the sample somewhat towards users who are curious or interested in the subject of this research. Participants were motivated to fill out the questionnaire in order to learn about the mechanisms and the reasons behind the Auramirror, thus appealing to their motivation to find out more. After completing the questionnaire, respondents were given a one-page description outlining the methods used by Auramirror and the ultimate goals of attentive technology (Appendix B). The anonymity of the respondents was guaranteed.

The questionnaire is attached as Appendix A. The first set of questions gathered demographic information as well as data on how the participants used the system. The three primary goals of the survey were met by the following questions:

1. What do you think the auras (blue blobs) represent?
2. Did Auramirror show who you were looking at (by merging the auras)?
3. Did Auramirror show who you were talking to?

The first question measures whether or not people found the visualization effective. The second question gauged whether people felt the algorithm was accurate, since the algorithm is meant to estimate gaze as an approximation for attention. The third question was used to find if people felt the algorithm effectively measured conversational attention. The first question was open-ended and the second and third question asked for opinions on a five-point scale between "Very Much" and "Not at all."[3]

5.2 THE EVALUATION

Half (55 percent) of the respondents were female and the average age was 26.8 but ranged from 8 to 64. Only 26 percent of the respondents used the mirror with someone that they did not previously know. Most (94 percent) indicated that they spent five minutes or less at the exhibit and almost a half (42 percent) spent less than two minutes. Three quarters (76 percent) said that they tried to play with or control the auras,

---

[3] The Ontario Science Centre staff who helped make this study possible insisted that the standard labels from "Strongly Agree" to "Strongly Disagree" had historically poor success with the audience at the Centre. To improve the rate of response I was advised to replace these words with the ones described.

for example by trying to pass their aura to someone else or to escape from under it. Half (48 percent) of the respondents visited the exhibit in a pair, but group sizes ranged from 1 to 14 people.

A large proportion of the respondents (75 percent) answered the question about what the auras represent. Their responses were post-coded into three categories: tracking, desired, and other. The *tracking* answers are responses that seem to answer the question of what the auras are tracking rather than what they represent. The *desired* responses mentioned attention, communication or interaction. All of the remaining responses were labelled *other* and the results are summarized in Table 1.

Table 1. Answers to the question "What do you think the auras represent?"

Category	Percent of Responses Analysed	Examples
Tracking	48% (44)	• Faces • Possibly sound, or heat • sound waves
Desired	32% (29)	• social bubble • an interaction or - lack of it - between two or more people • our focus of attention our thought focus
Other	19% (18)	• germ spreading • bactera • thinking

Table 2 shows the responses of the participants to the other questions. A chi-square test was done to compare the response distributions against a uniform distribution. For question 3 positive and negative responses were aggregated to remove the zero cell for the "strong disagreement" category. As expected, all questions showed distributions significantly different from the uniform distribution ($p < 0.001$). Questions 1 and 2 pertain to the primary goals of the survey and show that 65 and 69 percent either agree or strongly agree with the statements regarding the accuracy of Auramirror.

Table 2. Responses to selected questions

(percent answering in each category)

Question	Strong Agreement	Agree	Neutral	Disagree	Strong Dis-agreement	N =
1	10	55	22	12	1	121
2	8	61	21	9	1	120
3	26	62	8	4	0	120
4	12	49	10	26	3	120
5	3	7	15	62	14	120
6	8	21	21	42	8	120

Questions:

1. Did Auramirror show who you were looking at (by merging the auras)?
2. Did Auramirror show who you were talking to?
3. Did you have fun playing with Auramirror?
4. Did you feel the computer system was watching you?
5. If yes, did this make you feel uncomfortable?
6. If the computer system had recorded you, would this make you feel uncomfortable?

5.3 DISCUSSION

The main purpose of the survey was to describe the disposition of the participants towards Auramirror as a visualization tool. The system was programmed to estimate head orientation to determine who people were looking at, and 65 percent of the respondents agreed that the auras showed whom they were looking at. This suggests that most people found the attention-measuring algorithm accurate. Moreover, 69 percent agreed that the system showed the person that the participant was talking to. This suggests that most people thought that the algorithm was a good measure of conversational attention. With the limited exposure that people had to the system, it appears to have successfully represented attention.

Without being led in any way, 32 percent of the respondents described the auras as representing some form of attention, offering support for the notion that the auras are a good representation of attention beyond a quantitative expression of the measurements. If we consider tracking answers a misreading of the system's intent, then 62 percent of the remaining answers would support this claim.

Secondary results show that 62 percent of the people did feel that the system was watching them, but, surprisingly, 76 percent indicate that this did not make them feel uncomfortable. More surprisingly, half (50 percent) said that they would not feel uncomfortable if they were recorded, while 29 percent expressed that they would feel uncomfortable. Auramirror was presented as an exhibit at a science museum that promotes interaction, which may account for the benign perception of the system. Eighty-eight percent of participants agreed that participating in the exhibit was "fun". This suggests that is was successfully presented as a tool for play and that the system did not instigate a sense of discomfort. However, one of the hosts reported that upon reading the questions regarding whether or not they felt watched, many participants would quickly ask if there were other cameras in the Science Centre watching them [43].

The results show positive support for the goals set out for this survey. It should be noted, however, that this context for Auramirror is different from one where it might be used as a tool to inform design decisions. This two-person prototype is inadequate for capturing more subtle forms of attention between two users. Other more complex methods of measuring attention, for example using multiple video input sources, are suggested for future development to enhance the system. This prototype was used as an evaluation tool for itself. A next step to evaluating Auramirror would be to test hypotheses related to specific and demanding tasks, and to compare results against those obtained using alternative visualization methods.

6
CONCLUSION AND FUTURE WORK

6.1 OVERVIEW

This thesis examined the ways by which computers can recognize, monitor and display the foci of attention among people within a defined *attentive space*. It developed a prototype system to record the negotiation of attention or display it to the participants within the *attentive space*. The system elicits the feedback from participants or researchers when evaluating attention-tracking algorithms. Finally, the prototype was deployed at the Ontario Science Centre in Toronto where over fifteen thousand people [29] participated in the demonstration over the past four months. A user evaluation was carried out using survey research techniques.

The thesis placed the problem of attention recognition in the context of Human-Computer Interaction and discussed some of the precedents in terms of interfaces that use attention to improve interaction. This work also drew on examples using art to inform the design process and broaden the possibilities for the links between people and machines. It set the background for the development of the prototype mirror that reflects the users and superimposes on their images *attention auras* representing potential attention, and *attention tunnels* representing mutual attention between individuals.

The prototype was developed by modifying and integrating a number of available algorithms as well as creating some new ones. The software used OpenCV and the

DirectShow SDK. A boosted classifier was used to find faces in a captured image from the system's video camera, and a Kalman filter was used to track the found people. The aura showing potential attention was modeled as an elastic, spherical, triangular mesh that was created as an icosahedron-based geodesic sphere.

The system was deployed at the Ontario Science Centre in Toronto for over a month and may now become a permanent exhibit demonstrating computer perception. An evaluation questionnaire was completed by 121 participants. The assessment suggests that Auramirror does accurately estimate attention shared between pairs of people. Suggestions for improvements and enhancements to Auramirror follow.

6.2 FUTURE WORK

Auramirror may be extended in many innovative ways. The technology used by Auramirror can be used whole or in part to help elucidate attention and group interactions. Once the validity of a specific implementation with a particular attention measuring algorithm is established, we can explore ways to expand the system to recognize a greater range of expression. Finally, cosmetic improvements to the system are suggested.

6.2.1 Research to Validate Auramirror

Auramirror can be used as a tool to help build other prototypes of attention aware systems, and it can be used as a tool for performing experiments on human attention. Chapter 5 offered evidence that Auramirror can successfully visualize attention in the context of a museum exhibit. It remains to be shown that Auramirror can be used as a tool to improve the ability of researchers and designers to incorporate user attention when studying or creating user interfaces. This depends in part on the accuracy of the method of recognizing and monitoring attention, but more significantly on the user's ability to discern this accuracy from the system's visualization components. For example, if Auramirror fails to accurately measure attention and also conveys this failure to the participants who provide feedback to the researchers, then the algorithms for measuring attention that were incorporated in this version of Auramirror are flawed. Nevertheless, the Auramirror system would still be considered a success because it identified a faulty method for recognizing or tracking attention between people and helped point to the reason for the erroneous diagnostics. This information would allow researchers to fix errors early, before they become embedded in the later stages of the interface design and end up corrupting the project. To fulfill this role, the next step in the development of Auramirror would incorporate its evaluation as a potential research tool. This step might involve an experiment that requires the use of Auramirror in tasks that need to measure attention. Auramirror can be evaluated by determining how much it helps interface developers make good design decisions. In this way Auramirror could parallel Garabet's attempts to use performance art as an early evaluation tool.

6.2.2 Extensions of the Auramirror System

The auras depicted in Auramirror can be designed to be more informative. The addition of colour differentiation is one obvious dimension that has not yet been explored. Also, the constants that affect the motion dynamics of the auras can be easily altered in real-time to affect apparent viscosity and density. This is a more subtle change that can contribute another dimension to the representation of attention interfaces. Colour can be used to communicate emotions that have been detected from speech volume fluctuations, skin temperature, or pupil dilation. Another extension of the current prototype would record and reflect the duration of a relationship (shared attention) by using modulations in the opacity, luminosity or apparent viscosity of the auras. If the system is coupled with a method for identifying who the participants are, a cumulative duration spanning multiple uses of the system can be measured and the relative importance of relationships among individuals could be assessed over time. The sex or social status of an individual could be estimated using computer vision and represented by the colour or shape of their aura. In a scenario with people belonging to different organizations and with the means to identify this, individuals can be visually classified by applying different colours or patterns to their auras. This could make Auramirror a useful tool in the study of organizations or in research on the effect of sex or status differences in interpersonal communication. An obvious use would be in focus groups to help see how opinions are formed and communicated within the group. A variety of data on the individuals who enter the attentive space of Auramirror can be mapped into a range of texture, shape and dynamics.

Auramirror could also be used to indicate other forms of attention. If appropriate sensing capabilities were introduced, Auramirror could be made to visualize mediated attention, the foci of attention transmitted through communication technology. Measuring activities on instant messaging programs, phones, e-mail, and online community message boards, for example, would provide data indicating who is digitally attending to whom. These changes would enhance the Auramirror system by embedding more meanings into the images that are displayed in the mirror. The underlying artistic and functional messages would not be changed; rather, the increased density of the information communicated by the system would enhance it. The types of changes described above would make the system more personable. They could also make the system more invasive, perhaps more natural, and certainly more interesting.

6.2.3 Enhancements to the Existing Auramirror System

Auramirror was used as a prototype to examine its current attention-measuring algorithm. This examination has informed some of the decisions made in constructing the system and has indicated what future changes should be made. It is clear that the current two-person implementation does not fully capture the subtler aspects of interaction between two people, and the multiple camera setup described in Chapter 4 is a partial response to this problem. Auramirror provides only limited ability to indicate the attention of a person when his or her attention is not being returned by another person. This shortcoming can be overcome by using more precise head orientation data. In addition, the system could identify inanimate objects that the person is paying attention to and to display this information.

Aside from further work to evaluate the accuracy and to assess the potential scientific uses of Auramirror, more graphic design can enhance the reflected auras. A more artistic approach could offer a different rendering of the auras to make them more ethereal, or create an alternate, purely artistic display system that retains abstract auras but is more stylized in its use of the video data; for example, instead of a mirror image the auras could be texture-mapped with only captured audience members' faces. These changes would not alter the themes and concepts of Auramirror, only the presentation; they may entail a reduction in clarity of what is being viewed. This is what makes these changes undesirable for the more functionally oriented, current version of Auramirror. The reduction in clarity, however, could offer a partial response to privacy concerns.

6.3 CONCLUSION

Today's ubiquitous computers often interrupt face-to-face conversation by requesting attention at inappropriate times. The work presented here introduced Auramirror, an artistic installation that brings this issue to light. The Auramirror system uses the available power of today's processors to passively gather input data from a video camera to estimate and then visualize attention. This system was evaluated as an exhibit at the Ontario Science Centre by conducting a survey to determine people's reactions to the system. The results of the survey indicate that people felt the system accurately measured and visualized their conversational attention, within the context it was presented.

Auramirror is a video mirror that renders the virtual windows of attention through which we communicate with other people. Auramirror enlists fluid auras to convey the attention people pay to each other by recognizing and monitoring estimated mutual head orientation with the attentive space. It permits users to become more aware of the function of their attention in group interactions, and of the effects of disruption on this process. Artistically, Auramirror is designed to promote discussion about attention. It demonstrates that this information can be measured, albeit with limitations. Auramirror aims to help in the design process of interfaces that use attention.

Auramirror has successfully shown that attention can be measured and visualized by a computer system. It has shown that the estimation of gaze through head orientation is adequate for tracking attention in a simple two person scenario such as that found at the Ontario Science Centre exhibit. The exhibit also offers some support that the attention auras match what people expect attention might look like if it were made visible. Auramirror is presented as a successful proof of concept for a method of tracking and visualizing attention among groups of people. Auramirror also demonstrates one way that the pursuits of design and art can coincide to create a more interactive system that also provokes contemplation of design issues.

References

1. Abowd, G. D., et al. "Context-awareness in wearable and ubiquitous computing." *Proceedings of the 1st International Symposium on Wearable Computers ISWC* (1997): 179-180.

2. Argyle, M., and M. Cook. *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press, 1976.

3. Argyle, M., and J. Dean. "Eye-Contact, Distance and Affiliation." *Sociometry* 28.3 (1965): 289-304.

4. Bodow, Steve. "Wooden Expression." *Wired* Dec. 1999.

5. Bush, Vannevar. "As We May Think." *Atlantic Monthly* July 1945: 101-108.

6. Buxton, W. "Less is More (More or Less)." *The Invisible Future: The seamless integration of technology in everyday life.* Ed. P. Denning. New York: McGraw Hill, 2001. 145 – 179.

7. Chen, M. "Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconferencing." *Proceedings of CHI 2002 Conference on Human Factors in Computing Systems* (2002): 49-56.

8. CHI 2003. 19 Nov. 2003. <http://www.chi2003.org/index.cgi>

9. Comaniciu, D., V. Ramesh, and P. Meer. "Real-time tracking of non-rigid objects using mean shift." *Proceedings of IEEE Computer Vision and Pattern Recognition* 2 (2000): 142-149.

10. Darrell, T., et al. "A Virtual Mirror Interface using Real-time Robust Face Tracking." *Proceedings of Face and Gesture Recognition* (1998).

11. Dunne, A., and W. Gaver. "The Pillow: Artist-designers in the digital age." *Proceedings of CHI'97 Companion* (1997).

12. Engelbart, D. C. "A Conceptual Framework for the Augmentation of Man's Intellect." *Vistas in Information Handling Vol. 1.* Eds. Howerton & Weeks. Washington D.C.: Spartan Books, 1963. 1-29.

13. Fish, R. S., et al. "The VideoWindow system in informal communication." *Proceedings of Computer Supported Cooperative Work* (1990): 1-11.

14. Freund, Y., and R.E. Schapire. "Experiments with a new boosting algorithm." *Proceedings of the 13th International Conference on Machine Learning* (1996): 148-156.

15. Fukunaga, K., and L.D. Hostler. "The estimation of the gradient of a density function, with application in pattern recognition." *IEEE Transactions on Information Theory* IT-21 (1975): 32-40.

16. Garabet, Angela, Steve Mann, and James Fung. "Exploring design through wearable computing art(ifacts)." *Extended Abstracts of CHI 2002 Conference on Human Factors in Computing Systems* (2002): 634 – 635.

17. Gaver, W., A. Dunne, and E. Pacenti. "Projected Realities: Conceptual Design for Cultural Effect." *Proceedings of CHI'99 Conference on Human Factors in Computing Systems*. (1999): 600-608.

18. Gibson, J. *The Ecological Approach to Visual Perception*. New York: Houghton Mifflin, 1979.

19. Glenstrup, Arne John, Theo Engell-Nielsen. "Eye Controlled Media: Present and Future State." Bach. Thesis, Institute of Computer Science, University of Copenhagen, 1995. 18 Nov 2003 < http://www.diku.dk/~panic/eyegaze/article.html>

20. Greenspan, M. A., and P. Boulanger. "Efficient and Reliable Template Set Matching for 3D Object Recognition." *Proceedings of 3DIM99 the 2nd International Conference on 3-D Digital Imaging and Modeling* (1999): 230-239.

21. Hansen, J. P., and M. Støvring. "Udfordringer er blikfang." *Hrymfaxe* 3 (1988): 28-31.

22. Horvitz, E., A. Jacobs, and D. Hovel. "Attention-Sensitive Alerting." *Proceedings of UAI '99 Conference on Uncertainty and Artificial Intelligence* (1999): 305-313.

23. Jacob, R. J. "Eye tracking in advanced interface design." Eds. W. Baroeld and T. Furness. *Advanced Interface Design and Virtual Environments*. Oxford: Oxford University Press, 1995. 258-288.

24. Johnson, Steven. *Interface Culture*. San Francisco: Harper Edge, 1997.

25. Kahneman, D. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1973.

26. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." *Transaction of the ASME, Journal of Basic Engineering* 82D (1960): 35-45.

27. Kenner, H. *Geodesic Math and How to Use It*. Berkley: University of California Press, 1976.

28. Klasnja, Ana. Personal correspondence, 20 Nov. 2003.

29. Licklider, J. C. R. "Man-Computer Symbiosis." *IRE Transactions on Human Factors in Electronics* HFE-1 (1960): 4-11.

30. Licklider, J. C. R. "The Computer as a Communication Device" *Science and Technology* April 1968.

31. Lienhart, R., and J. Maydt. "An Extended Set of Haar-like Features for Rapid Object Detection." *Proceedings of IEEE International Conference on Image Processing* 1 (2002): 900-903.

32. Maglio, Paul P., et al. "Gaze and Speech in Attentive User Interfaces" *Proceedings of The third International Conference on Multimodal Interfaces* (2000): 1-7.

33. Marinetti, F. T. "Technical Manifesto of Futurist Literature" May 11, 1912.

34. Microsoft DirectX. 18 Nov, 2003. <http://msdn.microsoft.com>

35. Nielsen, J. "Non Command User Interfaces." *Communications of the ACM* 36.4 (1993): 83-99.

36. Norman, D. A. *The Invisible Computer*. Cambridge, MA: MIT Press, 1998.

37. Norman, D. A. *The Psychology of Everyday Things*. New York: Basic Books, 1988.

38. Open Source Computer Vision Library (OpenCV). 18 Nov. 2003. <http://www.intel.com/research/mrl/research/opencv>

39. OpenGL. 18 Nov. 2003. <http://www.opengl.org>

40. Osborn., B. "Futurism and the Futurists." 18 Nov. 2003 <http://www.futurism.org.uk/>

41. Preece, Jennifer, Yvonne Rodgers, and Helen Sharp. *Interaction Design – Beyond Human-Computer Interaction*. Hoboken, NJ: John Wiley and Sons, Inc., 2002.

42. Ramcharan, Vishnu. Personal communication, Aug. 15, 2003.

43. Redström, J., T. Skog, and L. Hallnäs. "Informative Art: Using Amplified Artworks as Information Displays." *Proceedings of DARE 2000 Designing Augmented Reality Environments* (2000): 103-114.

44. Schmidt, A., and K. Van Laerhoven. "How to Build Smart Appliances." *IEEE Personal Communications* 8.4 (2001): 66-71.

45. Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 3rd Ed. Reading, MA: Addison-Wesley, 1998.

46. Shackel, B. "Human Factors and Usability." Eds. J. Preece, and L. Keller. *Human-Computer Interaction: Selected Readings*. Upper Saddle River, NJ: Prentice Hall, 1990.

47. Shell, J., et al. "Interacting with Groups of Computers." *Communications of the ACM* 46.3 (2003): 40-46.

48. Shell, J., R. Vertegaal, and A. Skaburskis. "EyePliances: Attention-Seeking Devices that Respond to Visual Attention." *Extended Abstracts of CHI 2003 Conference on Human Factors in Computing Systems* 2003: 770-771.

49. Short, J., et al. *The Social Psychology of Telecommunications*. London: Wiley, 1976.

50. Skaburskis, A. W., et al. "AuraMirror: Artistically Visualizing Attention." *Extended Abstracts of CHI 2003 Conference on Human Factors in Computing Systems* (2003): 946-947.

51. Skaburskis, A. W., R. Vertegaal, and J. Shell "Auramirror: Reflections on Attention." To appear in *Proceedings of ETRA 2004 Symposium on Eye Tracking Research and Applications* (2004).

52. Stiefelhagen, R., and J. Zhu. “Head Orientation and Gaze Direction in Meetings.” *Extended Abstracts of CHI 2002 Conference on Human Factors in Computing Systems* (2002): 858-859.

53. Theeuwes, J. “Visual selective attention: A theoretical analysis.” *Acta Psychologica* 83 (1993): 93-154.

54. Vertegaal, R. “Attentive User Interfaces.” *Communications of ACM* 46.3 (2003): 30-33.

55. Vertegaal, R., et al. "Designing Attentive Cell Phones Using Wearable EyeContact Sensors." *Extended Abstracts of CHI 2002 Conference on Human Factors in Computing Systems* (2002): 946-947.

56. Vertegaal, R., et al. "Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes." *Proceedings of CHI 2001 Conference on Human Factors in Computing Systems*. (2001): 301-308.

57. Vertegaal, R., et al. "GAZE-2: Conveying Eye Contact in Group Video Conferencing Using Eye-Controlled Camera Direction." *Proceedings of CHI 2003 Conference on Human Factors in Computing Systems* (2003): 521-528.

58. Vertegaal, R. and Y. Ding. "Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronization?" *Proceedings of CSCW 2002 Conference on Computer Supported Collaborative Work* (2002): 41-48.

59. Viola, P., and M. Jones. “Rapid object detection using a boosted cascade of simple features.” *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition* 1 (2001): I-511-I-518.

60. Weiser, Mark. “The Computer for the 21st Century.” *Scientific American* September 1991: 94-104

61. Weiser, Mark, John Seely Brown. “The Coming Age of Calm Technology” PowerGrid Journal, v 1.01 (July 1996): 18 Nov. 2003 < http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm>

62. Welch, Greg, and Gary Bishop. “An Introduction to the Kalman Filter” TR 95-041, University of North Carolina at Chapel Hill. 18 Nov. 2003. <http://www.cs.unc.edu/~welch/kalman/kalmanIntro.html >

63. Wooding, D. S. “Fixation Maps: Quantifying Eye-Movement Traces.” *Proceedings of ETRA 2002 Eye-Tracking Research and Applications*. (2002): 31-36.

64. WebMuseum, Paris. 25 Jan. 2004. <http://www.ibiblio.org/wm>

65. Wren, C. R., et al. “Perceptive Spaces for Performance and Entertainment.” *Applied Artificial Intelligence* 11.4 (1997).

66. Wren, C., et al. “Pfinder: Real-Time Tracking of the Human Body.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.7 (1997): 780-785.

67. Yarbus, Alfred L. *Eye Movements and Vision*. Trans. Basil Haigh, Trans. ed. Lorrin A. Riggs. New York: Plenum Press, 1967.

Appendix A
Auramirror Questionnaire

Anonymous Auramirror Questionnaire

We very much appreciate you spending a few minutes and participating in a scientific experiment.

How many times have you visited this exhibit? ______

What is your age? ______ Sex? Male ◯ Female ◯

Please answer the following questions according to your latest visit to Auramirror.

How many people were with you in front of the mirror including you? ______

Did you use the mirror with someone you did not previously know? Yes ◯ No ◯

Approximately how much time did you spend at this exhibit?
Less that 2 minutes ◯ 2 to 5 minutes ◯ 10 minutes ◯ More than 10 minutes ◯

What do you think the auras (blue blobs) represent? ______________________________

____

Did you try to play with the aura (blob) or to control it? (For example by trying to escape from under it or pass it to someone else) Yes ◯ No ◯

Did Auramirror show who you were looking at (by merging the auras)?
Very Much ◯ YES ◯ I'm not sure ◯ NO ◯ Not at all ◯

Did Auramirror show who you were talking to?
Very Much ◯ YES ◯ I'm not sure ◯ NO ◯ Not at all ◯

Did you have fun playing with Auramirror?
Very Much ◯ YES ◯ I'm not sure ◯ NO ◯ Not at all ◯

Did you feel the computer system was watching you?
Very Much ◯ YES ◯ I'm not sure ◯ NO ◯ Not at all ◯

If yes, did this make you feel uncomfortable?
Very Much ◯ YES ◯ I'm not sure ◯ NO ◯ Not at all ◯

If the computer system had recorded you, would this make you feel uncomfortable?
Very Much ◯ YES ◯ I'm not sure ◯ NO ◯ Not at all ◯

Please share any further comments you may have with us. ______________________________

____

____ *(Continue on back if needed.)*

The author of Auramirror, Alexander Skaburskis, from the Human Media Laboratory at Queen's University School of Computer Science and The Ontario Science Centre thank you for your participation.

Appendix B
Explanatory Text Distributed to Participants of the Auramirror Survey

WHAT'S THE AURAMIRROR ALL ABOUT?

The AuraMirror is a digital media art piece, but it's also a new kind of computer interface and potentially, an interesting new scientific tool.

The AuraMirror recognizes human faces. When it finds your face, it creates a blob around it. Its computer then tries to determine where your attention is focused. If you interact with someone else, the computer will extend and merge your two blobs on the screen creating a "tunnel". But it's hard to see this yourself because the connection bursts as soon as you turn your head back to the screen. Why? The "tunnel" is a visual presentation of your attention towards the other person. When you turn your attention back to the screen, it's no longer on another person. In a funny way, this art piece conveys a clear point about the limitations of human visual perception: We can only attend to one thing at a time.

The blobs on the screen give us instant feedback about what the computer sees and recognizes to be our focus of attention. It's very easy to see when the computer is right or wrong. That makes it an effective new visualization and prototyping tool that could be applied to many other ways of measuring human attention.

Thanks to author Alexander Skaburskis, a researcher from Queen's University, for sharing his breakthrough work with us. Information collected here from questionnaires and computer logs will help him determine how well this concept works and how to improve on it. When the experiment is finished, information about the results will be published on our website. Check it out in couple of months.

Thank you to all Ontario Science Centre visitors contributing to the success of this experiment!

Ontario Science Centre
Toronto, 2003

For more info please call Ana Klasnja, Senior Multimedia Producer at 416 696 3269.

Vita

Name:	Alexander William Vischer Skaburskis
Place and Year of birth:	Vancouver, 1976
Education:	McGill University 1994-1999
	B.Sc. (Joint Honours Mathematics and Computer Science) 1999
	Queen's University, School of Computing 2001-2003
Experience:	Software Designer and Programmer, Hearthstone Alzheimer Care Ltd., Lexington MA. 1999-2001
	Teaching Assistant, Queen's University. 2001-2002
	Research Assistant, Queen's University, Human Media Lab. 2002-2003
Awards:	Queen's Graduate Award
Publications:	The Auramirror is currently being exhibited at the Ontario Science Centre. It will soon become a permanent exhibit there.
	Skaburskis, A. W., Vertegaal, R., Shell, J. S., "Experiences with Auramirror: Surveillance of Conversational Engagement" Submitted to *DIS 2004 Conference on Designing Interactive Systems*
	Skaburskis, A. W., Vertegaal, R., Shell, J. S., "Auramirror: Reflections on Attention" To appear in *Proceedings of ETRA 2004 Symposium on Eye Tracking Research and Applications* 2004.
	Shell, J. S., Vertegaal, R., Cheng, D., Skaburskis, A. W., Changuk Sohn, A. James Stewart, Omar Aoudeh, Connor Dickie, "ECSGlasses and EyePliances: Using Attention to Open Sociable Windows of Interaction"

Proceedings of ACM ETRA Symposium on Eye Tracking Research and Applications 2004.

Skaburskis, A. W., Shell, J.S., Vertegaal, R., and Dickie, C. "AuraMirror: Artistically Visualizing Attention." *Extended Abstracts of ACM CHI 2003 Conference on Human Factors in Computing Systems*, 2003.

Shell, J., Vertegaal, R, and Skaburskis, A. "EyePliances: Attention-Seeking Devices that Respond to Visual Attention." *Extended Abstracts of ACM CHI 2003 Conference on Human Factors in Computing Systems*, 2003.

Shell, J., Vertegaal, R., Mamuji, A., Pham, T., Sohn, C. and Skaburskis, A. "EyePliances and EyeReason: Using Attention to Drive Interactions with Ubiquitous Appliances" *Extended Abstracts of UIST*, 2003.