

Assessing Multiple Perspectives in Three Dimensional Virtual Worlds: Eye Tracking and All Views Qualitative Analysis (AVQA)

Sean P. Goggins
Drexel University
outdoors@acm.org

Jesus Guajardo
University of Missouri – Columbia
jesus.guajardo@health.missouri.edu

Matthew Schmidt
University of Missouri – Columbia
schmidtma@missouri.edu

Joi Moore
University of Missouri – Columbia
moorejoi@missouri.edu

ABSTRACT

The Human Computer Interaction community is beginning to wrestle with the challenges of effective interaction design for groups collaborating in 3 dimensional virtual worlds. Understanding of the requirements and design processes necessary to support these new technologies is still emerging. While the gaming community has developed heuristics for game design over a 25 year period, the adaptation of this knowledge for the design of virtual collaboration spaces has been limited. Following a research through design approach, we create novel integrations and applications of Human Computer Interaction research methods in an attempt to understand the collaborative work of a small group of individuals in a 3 dimensional virtual world. Specifically, the fine grained perspective provided by eye-tracking (ET), and a more coarse unit of analysis, provided by All-Views-Qualitative-Analysis (AVQA), are applied to understand non-gaming interaction in 3 dimensional virtual worlds. To formalize an integrated method, we describe three lenses for analyzing eye-tracking data in graphics intensive and animated virtual applications. Then, we combine those outcomes with a single method for integrating and analyzing the points-of-view of all subjects in a small group collaborating on a creative work task in a 3 dimensional virtual world. Through these examples, we illustrate the benefits of this type of research for 3 dimensional virtual world designers.

Author Keywords

Eye tracking, fovea, gaze, saccades, fixations, 3 dimensional Virtual Environments, evaluation, CSCW, CSCL.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., Human Computer Interaction): Miscellaneous.

INTRODUCTION

Over one billion hours of collaborative play have been logged in the game Halo [15]. Put into easily understood commercial terms, that's about 526,315 person years of work effort; or the entire population of Ljubljana, Slovenia

playing Halo instead of going to work, school or anywhere else during the day for an entire year. That milestone serves to underscore the increasing acceptance of virtual worlds as legitimate, compelling places for humans to interact with one another. This shift calls for recognition of virtual worlds as a construct for designing and managing collaborative work [23]. Qualitative studies of how gamers experience virtual worlds have explicated the details of preferred interaction styles [6], the formation of community [28] and provided generalized guidance for Human Computer Interaction designers [5,12,21,31] searching for new interaction design metaphors to support deeper engagement in collaborative work among geographically dispersed groups.

Dyck et al's [12] analyses of gaming as an interaction metaphor revealed the core dimensions of effortless community, learning by watching, deep customizability and fluid system-human interaction. These aspects of gaming environments hold promise for software designers seeking to build networks and communities among physically dispersed populations [31]. Gaming interaction design also challenges long accepted heuristics for interface design, including simplicity, consistency & ease of use for all users. These constructs are challenged by gaming's evidence that users enjoy learning a *game* quickly but then wish for it to become more difficult as they advance through different stages.

Interactions like those encountered in virtual worlds and games are more compelling than conventional windowing systems [26,29] and could form the basis for more engaging and productive Human Computer Interaction. The goal of our research is to explore methods for evaluating interaction design in virtual worlds, working toward the development of useful heuristics that could be applied by designers. We make a critical distinction between 3 dimensional virtual worlds and games: In the worlds we are exploring, collaboration is measured by creative output and group efficacy, not by a body count or other discrete score. The premise is that gaming has useful interaction design lessons for the Human Computer Interaction community to build upon, but that building worlds to support creative

collaboration is distinct from the challenge of creating a really great first person shooter (FPS) game. While the gaming community knows what works [7,33], design based research agendas for the development of collaborative 3 dimensional spaces are in a more formative state.

Interaction in 3 dimensional virtual worlds is more immersive, includes interactions with avatars and other social-emotional representations of people, and follows a set of interaction metaphors that are distinct from traditional user interfaces. Each environment typically embeds a distinct navigational scheme as well, though a limited number of conventions, like the WASD buttons on a QWERTY keyboard (used for walking around and setting direction of a user within the world), do exist. Avatars, in particular, add an emotional dimension to interaction that is not present in conventional user interfaces that support collaborative work or learning. In order to learn more about navigation and user experience in 3 dimensional spaces, we set out to define a protocol for capturing and analyzing the fine-grained details of what users pay attention to in virtual worlds, which is a noteworthy gap in the literature to date. We chose eye-tracking (ET) technology as one tool for understanding user attention because it captures information at the 1/16th of a second level of detail. While insightful, ET proved insufficient for building a comprehensive understanding of interaction in 3 dimensional virtual worlds. We chose eye-tracking technology over less specific usability evaluation methods like Morae alone because of the fine-grained detail provided. This led us to develop a complementary method for assessing 3 dimensional virtual world interactions through the integration of remote screen recording and formal usability lab observation of subjects, which we describe as the all-view-qualitative-analysis (AVQA) method.

EYE TRACKING TECHNOLOGY

Duchowski [11] reviewed the development and application of eye-tracking technology over the past 40 years, outlining historical phases in the fields of cognitive psychology, industrial engineering, neuroscience and computer science (Human Computer Interaction). The central value of eye-tracking is based on what Just and Carpenter [22] and others call the eye-mind hypothesis, which states that the location of a person's gaze directly corresponds to the most immediate thought in a person's mind. Early eye-tracking research was centered in cognitive psychology and neuroscience. Duchowski describes the present era in eye-tracking as *interactive*, with the most recent preceding era characterized as *diagnostic*. The key differentiation between these two eras is how eye-tracking technology is applied. In the diagnostic, eye tracking is focused on understanding what a user is viewing. In the interactive, eye-tracking technology becomes an aspect of user input. In both eras, eye-tracking application in the fields of industrial engineering and Human Computer Interaction are on the rise. In the context of interactions in windowed, 2D

spaces, the sequential framing of these eras is supported in the literature. However, the use of eye-tracking to refine virtual worlds for diagnostic purposes remains practically unexamined and potentially quite fruitful, particularly as a mechanism for understanding human-avatar interaction.

Capturing Eye-Tracking Data

Eye tracking takes two general forms. The first is monitoring of where the fovea (a part of the eye that can be tracked to indicate where a person is looking) is directed using special equipment in a laboratory setting and various means of making reasonable guesses at where a user is looking. Fovea tracking takes two primary forms: Those requiring a head mounted device, and those not requiring a head mounted device. There is significant expense associated with head mounted systems, and even more significant expense associated with non-head mounted systems, which have the advantage of capturing gaze information without subjecting the user to the attachment of a cyborg-like device to their head. The drawback of all fovea tracking systems in use today is that they require a laboratory environment, rendering the capture of gaze information in situ impractical.

Non-fovea-based eye-tracking systems have taken a variety of approaches, and most have been designed with the intention of overcoming either the cost or lab-based limitations of fovea-based eye-tracking systems. Mousetracks [1] captures mouse movement over an interface, with the expectation that users are following their on-screen cursor representation of the mouse movements with their eyes. Mousetracks implicitly extends the eye-mind hypothesis to the use of a pointing device in the interface: Where the mouse is pointing is where the user is looking; which in turn is what the user is thinking about. Tarasewich & Fillion [36] developed the enhanced restricted focus viewer, which achieves results similar to those of a full eye-tracking apparatus by blurring most of a user interface, rendering only a small section of the screen visible. This visible section is controllable by a user through a mouse or other pointing device.

Li, Babcock & Parkhurst [27] developed a head mounted eye-tracking apparatus from commodity hardware, in effect tackling the cost challenge of based-based eye tracking by open-sourcing the capture of gaze information. This accomplishment may signal the imminent commoditization of based-based eye-tracking technologies. It may also signal a decline in the exploration of non-based-based eye tracking, which has generally been less authentic or less precise.

Eye-tracking in the context of windowed applications is well suited for understanding how systems are learned, evaluating the efficiency of user interfaces when time is of the essence (emergency response applications, for example) and for interaction questions where traditional methods of studying usability, such as think aloud protocols, do not seem to reveal sufficient information regarding interface

limitations [24]. Similarly, we suggest that knowledge of how subjects in a virtual world learn and adapt to particular environments will be enhanced vis a vis the application of eye-tracking technologies to these interactions.

Three-dimensional environments do present a context where the We use avatars and images with limited movement to contrast fixation based and saccade based analysis with avatars.

Eye-Tracking Data Analysis

There are two primary views on how to analyze eye-tracking data, and how this information may be most useful for understanding user interactions with a computer system. The first view is that what are called fixations and saccades reveal the most compelling explanation of how a user is experiencing a system [25]. A fixation is a place where the user stares for a particular period of time. A saccade occurs when a user's gaze is moving from place to place in an interface. As might be expected, the distinction between fixation and saccade becomes more controversial with slower eye movements. The question "when does a fixation become a saccade?" has been the subject of debate in the eye-tracking research community. The second view of how to interpret and apply eye-tracking information is reviewing areas of interest (AOI) in the data. The AOI approach focuses on the density of total time spent in a particular area. The distinction between a fixation and a saccade is less critical than knowing where a user spends most of their time. Almost all analysis of eye tracking data falls into one of these two camps.

A number of studies [13,16,35] argue that analyzing areas of interest provides as much insight for many tasks as analysis that distinguishes between fixations and saccades. Other studies [38] find critical information present in the distinction between fixations and saccades, in some cases valuing the simpler, coarser but related concept of scanpaths. Given this controversy, as eye-tracking technology is applied to an increasing variety of Human Computer Interaction questions, ranging from web information usage [10] to visual search [3], it will not be surprising to discover eye tracking data analysis methods distinguished by the types of questions they help us to answer.

EYE TRACKING AND 3D WORLDS

Given the emergence of 3 dimensional virtual worlds, initiating work to understand interactions from the point of view of where users focus their gaze in these environments will aid in the development and refinement of these worlds for new purposes, including collaboration. One of the key distinctions between a virtual world and a windowed system is the presence of avatars and characters, which display emotional cues that are interpreted and acted upon by subjects. If these worlds are to be extended beyond applications focused on play, knowledge of the specific avatar characteristics attended to by subjects will aid

designers. Further, consideration of how collaboration, idea sharing and innovation can best be facilitated demands some attention to what the users of these systems pay attention to in general. These are fine-grained research questions that eye-tracking technology may help us answer.

Blascovich, Loomis, Beall, Swinth, Hoyt & Bailenson [2] suggest that anthropomorphized avatars – avatars that subjects earnestly believe represent another human being – have a higher degree of social presence than agent-avatars and as a result are more likely to induce authentic social influence effects in a collaborative virtual environment. There is an interaction between the nature or importance of this belief about the avatar's social presence, and the realism of the avatar's representation. In general, the social presence of the avatar is positively influenced if it is identifiably a "human-avatar" (the user or subject knows the avatar is controlled by another human). As a result, the behavioral realism of the on screen representation can be low and still result in an authentic social interaction. By contrast, if the avatar is perceived as being agent driven instead of representative of a human, the behavioral realism of the representation must be higher in order for social influence effects to occur.

Blascovich's theory suggests that low quality avatars may be beneficial if the subjects know that the avatars are representations of real people and that realistic avatars do not require the locus of control for the avatar to be "human". Therefore, improvements in the efficacy of virtual environments for fostering social interaction may be possible through either an increase in the identifiably human nature of the avatars, which can be established through human like dialogue or through an increase in the realism of the avatars representation of the human form. *Better graphics are not necessarily the answer to better 3 dimensional world collaboration.* Perhaps more significantly, if the attention of 3 dimensional world designers is turned toward collaboration and knowledge creation, fine-grained study of what subjects focus on in these contexts may reveal starkly different design characteristics for collaboration and knowledge creation than for gaming.

Human attention to others, and the choice of whom to and how to interact in the world is powerfully affected by emotions, and emotional expectations of users [37]. Facial features that people attend to when discerning the emotional state of others is well established in the literature. This knowledge serves as a backdrop for our development of a method for applying eye-tracking technology to the understanding of 3 dimensional virtual worlds, as our protocol is tested using a protocol that asks subjects to distinguish emotional states in human faces and avatars while eye-tracking data is captured. Subsequently, we integrate ideas from previous work like Mousetracks and the enhanced restricted focus viewer to explicate what users are attending to in a virtual environment with traditional usability lab protocols. The resulting session recording

data is then analyzed using a method we refer to as all-view-qualitative-analysis (AVQA).

The remainder of this paper is organized according to the manner in which we defined our protocol. First, two experiments where we captured eye-tracking information from subjects are described. Second, we describe the results from those experiments. Third, we discuss the results and provide a generalized method and toolset for consistently analyzing 3 dimensional virtual world interaction using heat maps. Fourth, we describe the AVQA method, and insights gathered from its early use. Last, we will articulate a research agenda for the continued development of technologies for understanding human-avatar interaction in 3 dimensional virtual worlds.

EXPERIMENT

Eye-Tracking apparatus

For the eye-tracking component of our method, subjects were seated in a desk chair in front of a computer screen with eye-tracking software configured. The eye tracking equipment used was an ASL R6 with VHT (video head tracking). Eye movements were recorded using this equipment, which was located in a human-computer interaction usability lab at a large university. The ASL R6 was selected over other eye-tracking technologies because it does not require a head mounted display.

Stimulus material

Several modes of virtual interaction were identified as stimulus because of their past use as images or 3 dimensional avatars with distinct emotional states. For this study, images used as stimulus material included pictures of real human faces and dynamic avatars that were obtained from <http://www.dotolearn.com>. To operationalize our protocol, a PowerPoint presentation was created displaying: a.) six faces of a man showing different emotional status (Figure 1), b.) six faces of a woman showing different emotional status, c.) an animated avatar exhibiting different emotions (Figure 2).

Protocol design

Subjects in this study are not provided with any incentives, time constraints or other external motivation for performance. Working with eye-tracking lab equipment in virtual environments like those we are studying are viewed as compelling by subjects, and recruitment was based on this intrinsic interest.

All subjects were provided with a description of the study context, and their role in it as outlined in the protocol. A researcher provided instructions and walked the subjects through an initial example of the first of two protocol components. The first protocol component was a set of static images that were numbered 1 to 6. Each of the pages had six individual faces with different emotions on each face. There were two sets of face-image pages, one with a

male face, and the other with a female face. The male version is shown below:

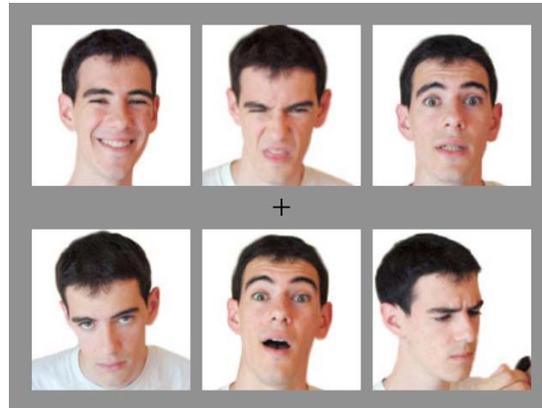


Figure 1 - Emotionally Expressive Faces

The second part of the protocol involved interacting with an animated avatar, and the subsequent identification of the appropriate emotion being displayed by the animated avatar (the avatar we used to display fear is shown in Figure 2). The limitations of the printed page prevent our inclusion of the animated version here, but it is available for download at: [{tbd in final version}](#)



Figure 2 - Animated, 3 dimensional Avatar Representing Fear

PHASE I RESULTS

Seven individuals participated in studies of our eye tracking protocol. The subjects included three females (one Caucasian, one African American and one Asian) and four males (three Caucasian and one Hispanic) ranging in age from 27 to 48. The first five yielded substantial preliminary data using a think-aloud protocol common to usability Studies [4]. The findings from these data enabled us to refine our methods for capturing further data using eye-tracking equipment. Having refined our protocols and analysis mechanisms, we performed extended eye tracking data analysis on the final two subjects' experiences with our virtual environment, yielding compelling results and a repeatable method.

It is noteworthy that two of the seven subjects had difficulty correctly identifying the emotions present in the static images, although all subjects correctly identified emotions represented by the animated avatar.

Results were gleaned from the analysis of eye tracking data captured while subjects went through the protocol. Data representing the X and Y coordinates of the subjects' gaze on a screen were obtained from the subjects using the ASL R6 eye tracking apparatus. A simple representation of the data in graph format (scatter plot) is below. What is being shown by these scatter plots is a single data point for each instance of gaze location captured by the equipment. High X, Y coordinates, for example, are representative of the subject looking in the upper right-hand corner of the screen, low X, Y coordinates are representative of the subject looking in the lower left-hand corner of the screen. Graphs were created from the two sessions for each subject. Below are examples from one session using the static image protocol for two different subjects.

The scanning pattern for these graphs is notable. One can quickly detect that each subject is scanning a series of equidistant objects, ultimately settling on one or two. This is notable because it represents an identifiable pattern of gaze behavior and provides a strong indication that the subjects' gaze information was accurately captured by the equipment.

The distinguishing nature of the gaze pattern when the subject is scanning static images is emphasized when compared to data from the animated avatar sessions, which show a clearly different scanning pattern in the case of subject B, and a nominally different scanning pattern in the case of subject A.

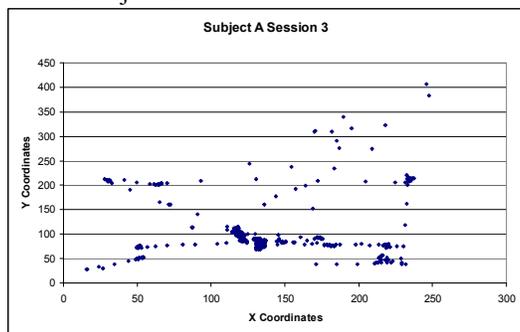


Figure 3 - Subject A, Avatar Data

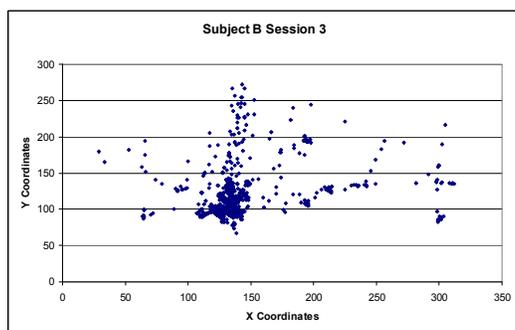


Figure 4 - Subject B, Avatar Data

Achieving our goal of integrating eye-tracking information with virtual environments required the addition of an image overlay to the representations of data from subjects A and B. This overlay brings the context of the subjects' gaze into focus. The overlays below were compiled by noting the X, Y coordinates of the image on the screen, calibrated for each subject, and then aligning the image with those coordinates. This was done by placing a scatter plot graphic over a photographic image used in the study. The two graphics were calibrated to be the same size, and coordinates 0,0 and 300,300 in perfect alignment, resulting in pixel-perfect accuracy, as shown in the overlays below:

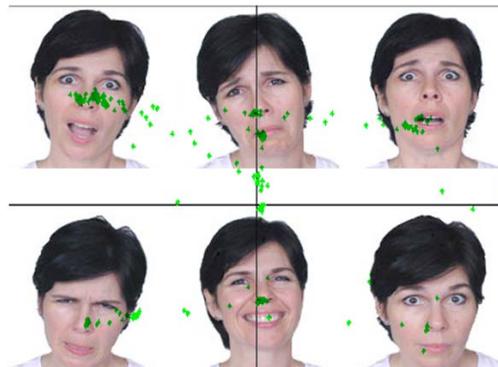


Figure 5 - Participant A, Static Image 2 Scatter plot Overlay

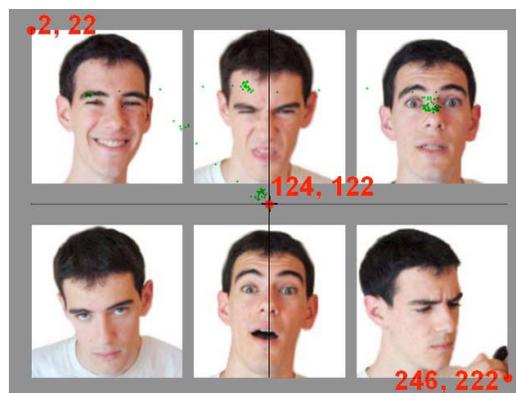


Figure 6 - Participant B, Static Image 1 Scatter plot Overlay

Reviewing the images side by side for each of the subjects analyzed in this manner reveals both similarity and difference in the gaze patterns among subjects: These initial overlays explicate which aspects of the virtual environment were attended to, but they do not provide the fine-grained cues needed to distinguish intensity of gaze in particular locations. To accomplish that, we analyzed the data using an open source heat mapping tool, initially intended for website usability, called "Clickheat" (<http://www.labsmedia.com/clickheat/index.html>). When we added the eye tracking data as heat map overlays to our

static images using 640x480 resolution, the heat maps were nearly perfectly aligned with the static images, with a skewing of the data in toward the center from the edges, which we were able to correct for in the following examples:

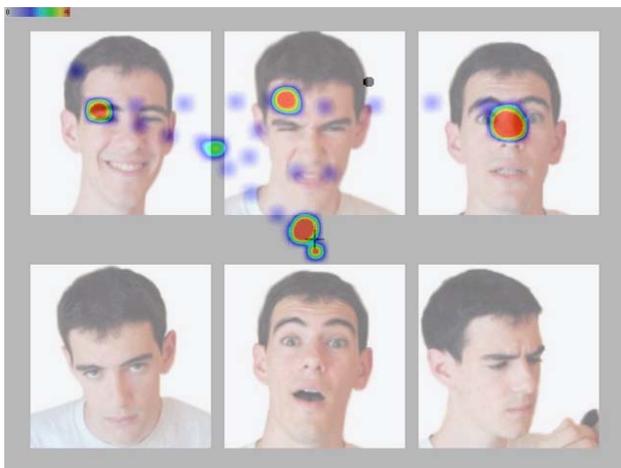


Figure 15 - Participant A, Static Heat Map 1

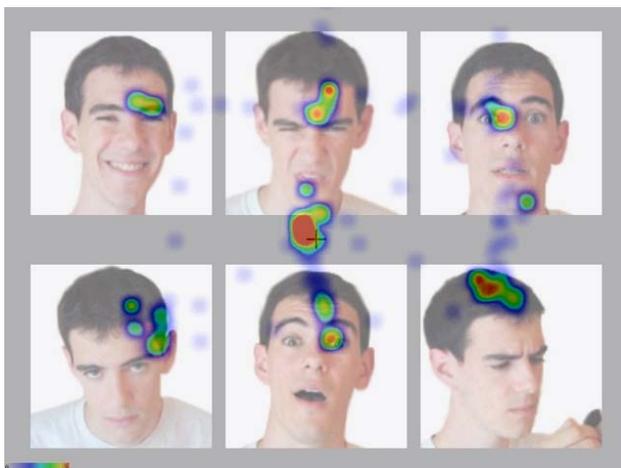


Figure 16 - Participant B, Static Heat Map 1

Eye-Tracking Over Moving Pictures (Avatars)

For the animated avatars, we conducted similar heat map analysis, and produced an animated data analysis that allows researcher's to follow the gaze in context. Below is a static image of that analysis tool:

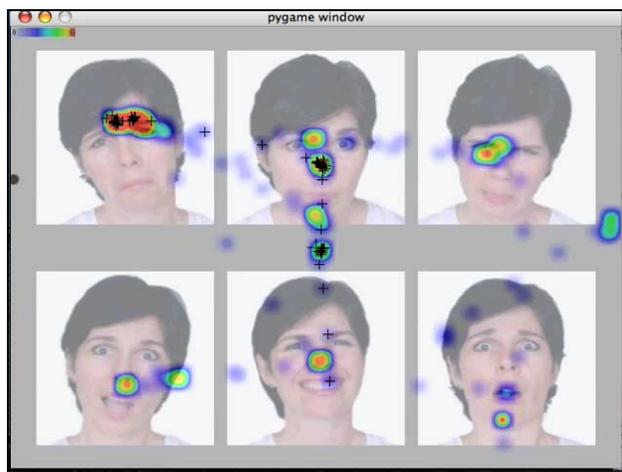


Figure 17 – Animated cursor following eye-tracking data points over time

Note the use of the heat map discussed in the static section, combined with animated black cross-hairs played for each 1/16th second of gaze from the user in Figure 17. By dynamically illustrating AOI and fixation-saccade data in this interface, we established a repeatable protocol for understanding gaze behavior in 3 dimensional virtual worlds. We later applied the exact same protocol, using the avatars illustrated earlier, to perform ET data analysis in a dynamic environment. The result is a repeatable process for understanding where users look when interacting with avatars in 3 dimensional virtual worlds.

METHOD FOR HEAT MAP ANALYSIS

Heat maps provide a powerful and intuitive way to graphically visualize gaze data. Typically, these visualizations are represented using the color red for drawing attention to areas that are intensely observed and the color blue for less-intensely observed areas. The longer a user gazes at a given area, the warmer the color in that area becomes. Areas that are not observed are represented using white (and sometimes black or even no color at all). To draw an analogy, just as peaks are represented in red and valleys in blue on topographical maps, so also is the intensity of gaze represented in heat maps.

Plotting eye tracking data on a graph enables the researcher to observe all eye tracking data points simultaneously, but does not provide a context for interpreting these data. Nonetheless, because these scatter plots are derived mathematically, the results of which can be replicated, the researcher can have confidence that these scatter plots provide a true representation of the underlying data. A simple way to provide context to these scatter plots is to overlay them onto the images that were used in the study; however, such images are not easily interpretable. The challenge for the researchers was to develop a method that would provide context, ease of interpretation, and would inspire the same degree of confidence as using a scatter plot. We surmised that if we would be able to

mathematically derive heat maps from our data with the same degree of replicability as scatter plots, then we would have met this challenge.

Principal to meeting this challenge was to pixel-perfectly align our heat maps to our mathematically proven scatter plots. Using linear transformations to convert the raw data provided by the eye tracking apparatus into X and Y coordinate data that corresponded to a resolution of 640x480 pixels, we were then able to feed these data into our heat map creation software. The initial output was skewed by 35 pixels, but we were able to correct for this by using linear transformations to resize the output. We then used the generated heat maps as semi-transparent overlays on the experiment images. Comparison of these heat maps with scatter plots demonstrated 1 to 1 correlation, the results of which were replicable. We consider this to be strong validation for the accuracy of our heat map overlays. Finally, the production of heat maps, with integrated tracking data over dynamic avatars provided a combination of AOI analysis and Saccade Fixation analysis required for more focused examination of 3 dimensional Virtual World interaction.

EYE-TRACKING ANALYSIS AND THE DAWN OF AVQA

The adaptation of eye-tracking technologies for the analysis of interaction in virtual worlds gave us compelling evidence of what individual users in these environments attend to. We built methods that are proving useful for identifying user interface affordances that are easily appropriated by novices in these environments. However, we found the data insufficient for understanding interactions *between* subjects in these 3 dimensional virtual environments.

In response to that gap in our findings, we developed a multi-threaded approach to analyzing user interaction in a 3 dimensional virtual space where a small group of users were coming together to complete a simple, collaborative task.

Studies of Novices in Virtual Worlds

To begin exploration of the question of how novices might adapt 3 dimensional virtual worlds for a collaborative work task, we studied groups using an online environment based on Croquet [8] while they performed a work task that involved the synthesis of multiple, related artifacts into a single, final version. We recorded the session from the point of view of three different subjects, observing one of the subjects in the usability lab at our university.

Our method included the installation of dynamic screen recording software on the local computers of two remote subjects. A third subject joined the group from our usability lab, which was equipped with a two monitor system, audio recording, facial recording, key stroke tracking and other features available in the usability evaluation software package, Morae. The user in our lab also started and used a laptop computer during the session. In addition to this group, we observed a fourth subject in a

different group who used the same 3 dimensional virtual worlds for collaboration from within our lab on a separate occasion.

This method is designed to capture user navigation and interaction from the point of view of each subject in a 3 dimensional virtual world. Our intention was to observe where users went and where they directed their avatars from the point of view of each user of the system. Like we did with the eye tracking data, we would then integrate and correlate these images with each other, observing in this case the same session from multiple points of view. This would be conceptually similar to observing a sporting event from multiple camera angles at the same time. Users did experience performance issues with the operation of a 3 dimensional virtual world and the stream recording of the whole session, however. As a result, we had to call off the recording after about 10 minutes in order to improve performance on two remote user computers. Later tests, using a MAC Pro and Dual Processor Dell Precision Workstation to observe users interacting with a virtual world on a local area network did result in proof a repeatable protocol. All sessions observed and recorded – three in total – lasted between 1.5 and 2 hours.

Though some of the lessons we learned about studying virtual worlds from multiple points of view were related to the technical infrastructure required for such study, we were also able to evolve and experiment with our ideas for analysis. In addition to the data captured via recording equipment, field notes were gathered during both laboratory sessions, and several subjects participated in follow up interviews and posted reflections to discussions boards in an online course community.

We analyzed field notes using a line by line open coding technique, and integrated one researcher's list in a set of common terms and observations. When initially introduced to the environment, users interacted with each other playfully, learning to jump, move around, and drag documents in and out of the environment. The chat tool, which was integrated in the environment we used, was the primary means of communication between subjects. An earlier pilot study which integrated voice discussion with interaction in the 3 dimensional virtual world proved to have too much latency, and take up too much bandwidth to be feasible for users who were not on the same local area network. This limitation is a technological one, and will likely be resolved in future releases of the environment.

Phase II Results

Our observations of interactions between users, undertaken both 'live' and from the side by side comparison of simultaneous session recordings, revealed four primary patterns:

- Group formation appears similar to that experienced in the real-world

- Novices bring social cues with them from the physical world
- 3 dimensional virtual world collaborators are usually engaged in multiple simultaneous activities
- Environment use declines following group formation.

Group Formation

Like groups placed together in a physical setting, the absence of clear boundaries and well understood organizational structure led to initial group formation ‘storming, forming, norming’ activity, although it appeared playful. This group formation sequence occurred quickly, especially by comparison with similar activities among physically collocated groups. This is consistent with past findings of game designers and researchers. The groups observed first experimented for about 15 minutes with the novelty of the environment. In the cases we observed, one individual then proceeded to begin working on the collaborative task that was the group’s objective, and other members followed. Analysis of the side by side recordings available show that most users paid attention to the artifacts on the walls in the virtual world, and then, only later, to the construction of a shared artifact. Chat was the dominant coordination technology during this period.

In interviews following the sessions one subject, Khan, reported that, although the group he was in could have completed their work without the virtual environment, he felt more comfortable and open with the group because he was able to interact with them visually, and ‘see’ what other members were doing.

Social Cues

Participants in collaborative activities in online virtual worlds seem to pay attention to social cues and norms common in the physical world. This impedes discovery, and the reading of shared artifacts in the 3 dimensional worlds we examined. For example, subjects would not move their avatars in front of or close to another avatar when reviewing shared material. Khan reported “I did not want to get in Steven’s way. So, I dragged the document to my desktop to read it, then copied my changes into the online version once Steven was done”. In this sequence, one of Khan’s course mates was reading a document on a wall, and Khan avoided ‘invading his personal space’, and chose instead a much more cumbersome and time consuming approach to integrating his ideas with those of the group.

Multiple Simultaneous Activities

In our lab, Khan used the monitor on the Morae equipped station, as well as his own laptop for participation in the group. Lucy, our other subject in the lab in a separately observed session, restricted her work to a single monitor, but nonetheless frequently referred to her notes, and “slipped out” of the 3 dimensional virtual world to check assignments and status in an online discussion board. The

world itself, engaged in from the lab computers, served as an instrument or mode of communication. Information seeking, note taking and assimilation of shared knowledge occurred either by jumping out of the 3 dimensional virtual environment completely, or by turning away from the environment to another computer, as in the case of Khan’s laptop.

The insight here is that the collaborative world provided by the software makers is not the totality of the knowledge creation and innovation for groups in these environments. Distinct from a gaming context, the subjects we have observed to date are *not immersed* in the 3 dimensional virtual worlds they use to collaborate. Creative activity, unlike gaming, appears to support the ‘in and out’ nature of these engagements.

Decline in Environment Use over Time

3 dimensional virtual worlds proved useful to our subjects as a social, “group formation” place at the beginning of a creative collaboration, but become less central to the joint endeavor as creative work progressed. Our observation across sessions was that the distributed teams built an identity and a shared understanding of the problem within the virtual world, but moved fluidly in and out of it as creative work (in this case the construction of a course design document) increased. In the case of Khan’s group, while they were willing subjects in the 3 dimensional virtual world and actually designed their collaborative work product to incorporate it as a central mechanism for learning (describing at one point their team ‘jumping together in collaborative ecstasy’), they never did return to it for collaboration after the initial creative session. Instead, they relied on easier to coordinate, asynchronous means for collaboration like discussion boards, shared editors (wikispaces, etc.) and archived chat.

The Emergence of a Method

From our experiences studying the engagement of an online group performing a creative task in a virtual environment, we developed a protocol that we intend to use for ongoing analysis of engagement in virtual worlds. The protocol includes the use of Camtasia studio for screen recording on Windows based PC’s (<http://www.camtasiastudio.com>) and Screen Record 1.6.1 (<http://mac.sofotex.com/download0121771.html>) for MAC based PC’s. By having users login to the system after starting the screen recording software, and making a clock available as the first shared object in the workspace, we have been able to synchronize session recordings and analyze relative position and perspective of interaction across multiple subjects in a single online session. Detailed, minute by minute annotation of each session is then compared across sessions, and specific types of interactions are noted and analyzed at varying depths, depending upon the research question. The four primary types of virtual engagement between avatars observed in our pilot study included:

1. Playful engagement, like jumping around.
2. Shared engagement around an artifact
3. Disengagement or isolation (walking around and exploring the space independently)
4. Chat engagement.

There were a very limited number of ‘face to face’ engagements between avatars in the collaborative space we studied. This could be a side effect of the task, or possibly a more generalizable discovery.

IMPLICATIONS FOR DESIGN

Multi-view and fine-grained views of user interactions with 3DVLE’s will provide next generation interaction designers with a complete view of how participants are adapting systems in use. At the level of interactions between participants in a virtual world, being able to see the point of view of each participant provides an opportunity to interpret whether or not the world, as designed, leads to social interaction or to the social interaction that is desired. Through eye-tracking, we gain insight into the specific behaviors and practices of individual participants. Together, these two methods show designers how the individual participants interact with the tools, and also how the individual participants interact through the tools. Previous, independent instantiations of these analysis are independently useful for usability evaluation. Together, these methods provide a complete evaluation toolkit that will enable designers to co-evolve the tools, and interactions within tools as they design next generation socio-technical systems.

CONCLUSION

This paper has presented a year of work developing two distinct but interconnected methods for understanding how subjects engage in 3 dimensional virtual worlds. Our goal has been to ensure that we understand the questions, and design considerations that will be faced by interaction designers who address the challenges associated with next generation Human Computer Interaction. We believe this will include the application of 3 dimensional virtual worlds and collaborative spaces. With that clarity, we searched for existing methods that could be applied to the study of 3 dimensional virtual environments, and produce insights that skilled designers could then use to develop heuristics for the design of 3 dimensional collaborative spaces.

Critically, we have recognized a distinction between 3 dimensional collaborative spaces and games. We also recognize and have sought to encourage the application of heuristics developed by the game industry, where they fit 3 dimensional virtual collaboration.

Our integrated approach has four main benefits to the Human Computer Interaction community. First, the application of existing eye-tracking technology to 3 dimensional virtual world navigational challenges enables the exposition of basic navigational issues experienced by novices. Second, by providing time-series feedback of where subjects are looking with respect to an avatar’s face in a 3 dimensional virtual world, questions about emotion,

emotional engagement and social presence cues are discoverable. Third, we have applied tangential concepts from the eye-tracking literature and proposed that, just as one may discern where a user is looking in a windowed interface based on where their mouse is pointing, researchers may also be able to garner direct insight into what a group of users is looking at in a 3 dimensional virtual world, simply by *observing multiple perspectives on the same virtual world session side by side*. Fourth, the integration of these methods to examine a series of different 3 dimensional virtual environments – ideally by a number of different research teams – will advance the use of 3 dimensional collaborative spaces for work and learning at a very rapid pace.

3 dimensional virtual world technology is presently very successful at engaging people in play. We hope that by sharing these methods, and suggesting their integrated use, we will inspire serious discussion about how to design and evaluate 3 dimensional virtual worlds for collaboration and learning across geographical, linguistic and cultural boundaries. We seek to advance the role that design based research, Human Computer Interaction and the design profession in general are able to play in the creation of better worlds within our world.

REFERENCES

- [1] Arroyo, E., Selker, T., & Willy, W. (2006). *Usability Tool for Analysis of Web Designs Using Mouse Tracks*. Paper presented at the CHI 2006.
- [2] Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2001). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*.
- [3] Byrne, M. D. (2001). ACT-R/PM and Menu Selection: Applying a Cognitive Architecture to Human Computer Interaction. *International Journal of Human-Computer Studies*, 55, 41-84.
- [4] Carroll, J. M. (2003). *Human Computer Interaction Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. San Francisco, CA: Morgan Kaufmann.
- [5] Chen, V. H.-h., Duh, H. B.-L., Kolko, B., Whang, L. S.-M., & Fu, M. C.-H. (2006). *Games in Asia Project*. Paper presented at the ACM CHI 2006.
- [6] Clarke, D. & Duimering, R. (2006). How Computer Gamers Experience the Game Situation: A Behavioral Study. *ACM Computers in Entertainment*, 4(3), 1-23.
- [7] Crawford, C. (1982). *The Art of Computer Game Design: Reflections of a Master Game Designer*. McGraw-Hill.
- [8] Croquet Consortium, T. Croquet: Invent the Future. Retrieved September 14, 2007, from http://www.opencroquet.org/index.php/Main_Page
- [9] Crowe, E. C. & Narayanan, N. H. (2000). *Comparing Interfaces Based on What Users Watch and Do*. Paper presented at the Eye Tracking Research And Applications Symposium 2000.

- [10] Cutrell, E. & Zhiwei, G. (2007). *What Are You Looking For? An Eye-Tracking Study of Information Usage in Web Search*. Paper presented at the CHI 2007, San Jose, CA.
- [11] Duchowski, A. (2002). A Breadth-First Survey of Eye-Tracking Applications. *Behavior Research Methods, Instruments and Computers*, 34(4), 455-470.
- [12] Dyck, J., Pinelle, D., Brown, B., & Gutwin, C. (2003). *Learning From Games: Human Computer Interaction Design Innovations in Entertainment Software*. Paper presented at the Graphics Interface Proceedings 2003, Halifax, Nova Scotia Canada.
- [13] Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). *Eye Tracking in Web Search Tasks: Design Implications*. Paper presented at the ETRA 2002, New Orleans, LA.
- [14] Greene, H. H. (2006). The Control of Fixation Duration in Visual Search. *Perception*, 35, 303-315.
- [15] Grossman, L. (2007). *The Man in the Mask. Technoculture*
- [16] Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). *The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye Tracking*. Paper presented at the chi 2006, Montreal, Quebec Canada.
- [17] Heminghouse, J. & Duchowski, A. (2006). *iComp: A Tool for Scanpath Visualization and Comparison*. Paper presented at the APGV 2006.
- [18] Hornof, A. J. & Halverson, T. (2002). Cleaning up Systematic Error in Eye-Tracking Data by Using Required Fixation Locations. *Behavior Research Methods, Instruments and Computers*, 34(4), 592-604.
- [19] Isaacowitz, D. M. (2006). Motivated Gaze: The View from the Gazer. *Current Directions in Psychological Science*, 15(2), 68-72.
- [20] Joachims, T., Ganka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. *ACM Transactions on Information Systems*, 25(2), 1-27.
- [21] Jorgensen, A. H. (2004). *Marrying Human Computer Interaction/Usability and Computer Games: A Preliminary Look*. Paper presented at the NordiCHI '04, Tampere, Finland.
- [22] Just, M. A. & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychology Review*, 87, 329-354.
- [23] Kaptelinin, V. & Czerwinski (Eds.). (2007). *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments*. Boston, MA: MIT Press.
- [24] Karn, K. S. (2006). *Eye Tracking for Usability testing, You've Got to Know Your Strengths and Weaknesses*. Paper presented at the CHI 2006.
- [25] Karn, K. S., Ellis, S., & Juliano, C. (1999). *The Hunt for Usability: Tracking Eye Movements*. Paper presented at the CHI 1999.
- [26] Larson, J. (2007). Out of the Video Arcade, into the Office: Where Computer Games Can Lead Productivity Software. *Interactions*, January/February, 18-22.
- [27] Li, D., Babcock, J., & Parkhurst, D. J. (2006). *openEyes: A Low-Cost Head-Mounted Eye-Tracking Solution*. Paper presented at the ETRA 2006.
- [28] Nardi, B. & Harris, J. (2006). *Strangers and Friends: Collaborative Play in World of Warcraft*. Paper presented at the CSCW '06, Banff, Alberta, Canada.
- [29] Paiva, A., Andersson, G., Hook, K., Mourao, D., Costa, M., & Martinho, C. (2002). SenToy in FantasyA: Designing an Affective Sympathetic Interface to a Computer Game. *Personal and Ubiquitous Computing*, 6, 378-389.
- [30] Pelz, J. B., Canosa, R. L., Kucharczyk, D., Babcock, J., Silver, A., & Konno, D. Portable Eyetracking: A Study of Natural Eye Movements.
- [31] Rapeepisarn, K., Wong, K. W., Fung, C. C., & Depickere, A. (2006). *Similarities and Differences Between "Learn Through Play" and "Edutainment"*. Paper presented at the Proceedings of the 3rd Australasian conference on Interactive entertainment IE '06.
- [32] Reeder, R. W., Pirolli, P., & Card, S. K. (2001). *WebEyeMapper and WebLogger: Tools for Analyzing Eye Tracking Data Collected in Web-Use Studies*. Paper presented at the CHI 2001.
- [33] Rollings, A. & Adams, E. (2003). *Andrew Rollings and Ernest Adams on Game Design*. New Riders.
- [34] Salvucci, D. D. & Goldberg, J. H. (2000). *Identifying Fixations and Saccades in Eye-Tracking Protocols*. Paper presented at the ETRA 2000.
- [35] Santella, A. & DeCarlo, D. (2004). *Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest*. Paper presented at the ETRA 2004.
- [36] Tarasewich, P. & Fillion, S. (2004). *Discount Eye Tracking: The Enhanced Restricted Focus Viewer*. Paper presented at the Tenth Americas Conference on Information Systems, New York, NY.
- [37] Wadlinger, H. A. & Isaacowitz, D. M. (2006). Positive Mood Broadens Visual Attention to Positive Stimuli. *Motiv Emot*, 30, 89-101.
- [38] West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). *eyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences*. Paper presented at the ETRA 2006, San Diego, CA.
- [39] Whittaker, S. (1996). *Talking to Strangers: An Evaluation of the Factors Affecting Electronic Collaboration*. Paper presented at the CSCW '96, Cambridge, MA.
- [40] Zhai, S., Morimoto, C., & Ihde, S. (1999). *Manual and Gaze Input Cascaded (MAGIC) Pointing*. Paper presented at the CHI 1999.
- [41] Nielsen, Jakob (2007). Video Retrieved from http://www.useit.com/alertbox/video_talkinghead_eyetrack.wmv, August 28, 2007.