

Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System

Susanna Nilsson^{*♦}, Torbjörn Gustafsson^{*} and Per Carleberg^{*}

[♦]Linköping University
(Sweden)

^{*}XMReality, Linköping
(Sweden)

ABSTRACT

Eye contact in human conversations is a natural source of information about the visual attention of people talking, and also indicates who is speaking to whom. Eye gaze can be used as an interaction method, but gaze tracking can also be used to monitor a user's eye movements and visual interest. This paper describes how gaze-based interaction can be used and implemented in an Augmented Reality (AR) system. The results of the preliminary tests of the gaze-controlled AR system show that the system does work, but that it needs considerable development and further user studies before it can be a realistic option in real end user settings.

Keywords: *Augmented Reality, gaze-controlled augmented reality, mixed reality, gaze-based interaction*

Paper Received 14/11/2008; received in revised form 04/04/2009; accepted 28/04/2009.

1. Introduction

Using eye gaze as input to technical systems is not a new concept. However, to a large extent, the methods of interacting with systems through eye gaze or eye gestures have been too strenuous, complicated or expensive for widespread use in the domain of human-computer interaction. Instead, eye gaze research has mainly developed as a tool for people with various limitations that make it difficult or impossible to interact with computers through traditional means such as keyboards, mouse or voice control.

Cite as:

Nilsson, S., Gustafsson, T., & Carleberg, P. (2009). Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System. *PsychNology Journal*, 7(2), 175 – 196. Retrieved [month] [day], [year], from www.psychology.org.

* Corresponding Author:

Susanna Nilsson
Linköping University, SE-58183 Linköping, Sweden
e-mail: susni@ida.liu.se

Eye contact in human conversations is a natural source of information about the visual attention of people talking, and also indicates who is speaking to whom.

This source of information can also be of use when communicating with a technological system. Eye gaze tracking can be used as part of an interaction method, but gaze tracking can also be used to monitor a user's eye movements and visual interest. This allows the system to recognize the possible intentions of the user quickly and thereby improve the user interface.

In Augmented Reality (AR) systems, real and virtual objects are merged and aligned in relation to a real environment, and presented in the field of view of a user. AR applications that give hierarchical instructions to users often require some feedback or acknowledgement from the user in order to move to the next step in the instructions. It should be possible to give this feedback quickly and without interrupting the ongoing task. Many different types of interaction techniques have been used in the domain of AR; there are numerous examples of systems that use manual input, gestures and/or speech interfaces (Billingham, Kato, & Poupayev, 2001; Gandy et al., 2005; Henrysson, Ollila, & Billingham, 2007). However, there are situations where speech and gesture may not be appropriate. For instance, during surgical procedures in an operating room, the surgeon may have difficulties manually interacting with technical devices because of the need to keep the hands sterile. Voice interaction with a system may also not be appropriate due to surrounding noise or filtering problems. There is one modality that can overcome the issues of noisy environments, keeping hands sterile and the need to work with both hands while at the same time trying to interact with a computer or an AR system, and that is the visual modality. The aim of this paper is to present an AR system with an integrated gaze tracker, allowing quick feedback from the user to the system, as well as an analysis of the user's gaze behaviour.

2. Mixed and Augmented Reality

Mixed Reality is a field of research as well as the collective name for the techniques aiming at combining and fusing virtual elements into the perceived real world. The term "mixed reality" aims to capture the conceptual ideas behind the technologies used – the blending, merging or mixing of different realities. Milgram and Kishino (1994) described the field of Mixed Reality (MR) as a continuum in which the different types of systems and applications can be placed (see Figure 1). In the far left of the continuum, systems

that aim to add only minimal virtual elements to the otherwise un-manipulated perceived real world can be found. These systems are known as Augmented Reality systems, where the aim is not to create an immersive virtual world (as opposed to Virtual Reality (VR) systems to the far right in the continuum), but rather augment the real world with added information or experiences. This paper focuses on an Augmented Reality (AR) system.

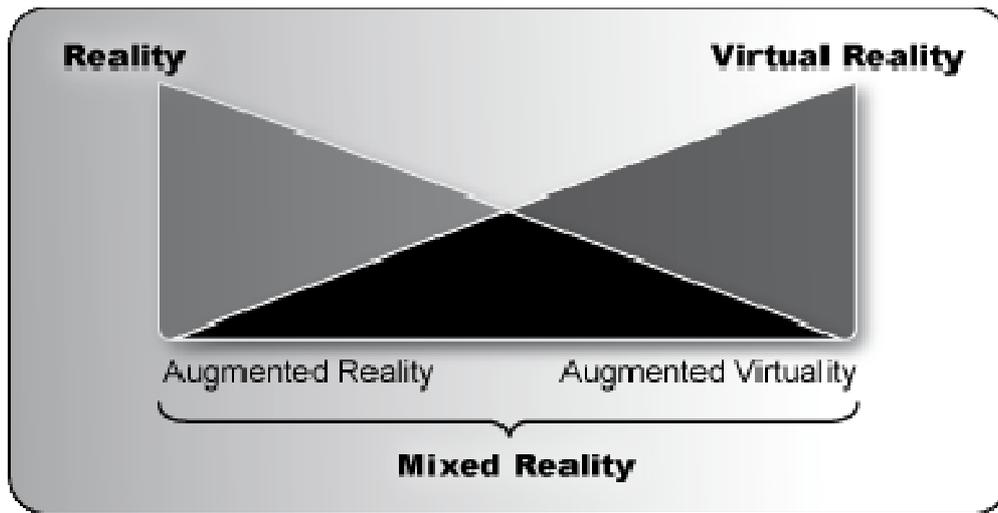


Figure 1. The Mixed Reality Continuum (after Milgram and Kishino, 1994).

Even though it may be an interesting question, this paper will not discuss the notion of "reality" or go into a philosophical debate about what is real and what is not. For the purpose of this text, "reality" is simply what humans perceive as their surrounding in their normal life. Given this definition of reality, "merging realities" could simply mean merging two images of the world together. Azuma (1997), however, mentions three criteria that have to be fulfilled for a system to be classified as an Augmented Reality system: It combines the real and the virtual, it is supposed to be interactive in real time (meaning that the user can interact with the system and get a response from it without delay), and it is registered and aligned in three dimensions. AR applications can be found in diverse domains, such as medicine, military applications, entertainment, technical support and industry applications, distance operation and geographic applications.

Designing usable and user-friendly interfaces is crucial when developing new systems. One area of research that has a close connection to usability and the end user community is research on interaction modalities. The user of an AR system can either be a passive recipient of information or an active part of an interaction process.

AR systems that use gestures and speech for interaction have been developed (Billinghurst, Kato, & Poupyrev, 2001; Gandy et al., 2005; Henrysson, Ollila, & Billinghurst, 2007). However, there are examples of situations where speech and gesture may not be appropriate. For instance, during surgical procedures in an operating room, the surgeon may have difficulties interacting with technical devices since both hands are usually occupied. Voice interaction with a system may not be appropriate due to surrounding noise or filtering problems. Another example can be found in a modern-day auto repair workshop, where a mechanic is doing repair work on an engine. Having greasy hands or being occupied with tools in both hands certainly would make interacting with a manual or technical support system difficult. There is one modality that can overcome both the issues of noisy environments and the need to work with both hands while at the same time trying to interact with a computer or an AR system, and that is the visual modality.

Many advances in direct manipulation, gaze-based interaction, pattern matching and recognition, gesture and natural language interaction have been made during the past decades. Apart from being used as an interaction method, gaze can also be used to monitor a user's eye movements and visual interest in usability studies and design evaluation. Gaze tracking allows the system to recognize the possible intentions of the user quickly, and this has been used to improve the interaction in multimodal user interfaces (Qvarfordt, 2004). Several of these results have also been implemented in the AR research area, but there are few examples of gaze-based interaction in this domain (Gustafsson & Carleberg, 2003; Gustafsson et al., 2005a; Sareika, 2005).

3. Technology Used in Augmented Reality Systems

Technically, there are two different solutions for merging reality and virtuality in real time today: video see-through and optical see-through. At first glance, the latter of these is the preferable solution, but it has some technical and practical difficulties (Azuma, 1997, 2001; Kiyokawa, 2007). For one, the virtual projection cannot completely obscure the real world image – the see-through display does not have the ability to block off incoming light to an extent that would allow for a non-transparent virtual object. This means that real objects will shine through the virtual objects, making them difficult to see clearly. The problem can be solved in theory, but the result is a system with a complex configuration. There are also some issues with the placement of

virtual images in relation to the surroundings in optical see-through displays. Since the virtual objects presented to the user are semi-transparent, they give no depth clues to the user. Instead, the virtual objects seem to lie in the same focal plane, whereas in natural vision, objects are perceived in different focal planes (Gustafsson et al., 2004; Haller, Billinghurst & Thomas, 2007).

A way to overcome the problems with optical see-through is by using a camera in front of the user's eyes, and then projecting the camera image onto a small display in front of the user's eyes (video see-through) (Azuma, 1997; Gustafsson et al., 2004; Kiyokawa, 2007). The virtual images are added to the real image before it is projected, which solves the optic see-through problem of surrounding light, as well as gives control over where the virtual objects are placed. This method, however, has other problems, such as the limited field of view, lack of peripheral displays and the slight offset caused by the fact that the camera's position can never be exactly where the eyes are located. This gives the user a somewhat distorted experience, since the visual viewpoint is perceived to be where the camera is (Azuma, 1997). The difference between the bodily perceived movement and the visual movement as seen through the display can have an effect on the user's experience of the system, in some cases even causing motion sickness (Stanney, 1995). Despite these problems, there are important advantages with a video see-through solution. One has already been pointed out – the ability to occlude real objects – and another is that the application designer has complete control over the presented image in real time since it is run through the computer before it is presented to the user. In the optical see-through design, only the user will see the final augmented image. To conclude: there is a trade-off between the optic see-through systems and the camera based video see-through systems, and the available resources often determine the choice of solution.

Regardless of what display solution has been chosen for an AR application, the most important issue to solve is how and where to place the virtually generated image. In order to place the virtual information correctly, the AR system needs to know where the user and user viewpoint is. This means that the system has to use some kind of tracking or registration of the surrounding environment. There are different techniques to do this, and several of them can be combined to ensure more reliable tracking of the environment (Haller, Billinghurst & Thomas, 2007). Tracking is normally done by using different sensors to register the surrounding environment. This sensor information is then used as a basis for placing the virtual information (Azuma et al., 2001). When using a video see-through technique, the AR system is already equipped with a visual

sensor – the camera – which allows vision based tracking. Although the development of marker-less techniques may dominate the future, the marker tracking technique has been one of the most commonly used techniques; it makes use of visual markers that can be recognized by feature tracking software (Kato & Billinghurst, 1999). Figure 2 below shows an example of a marker that can be used for this purpose.

The marker tracking technique used in the studies presented in this chapter is based on ARToolkit, which is an open software library for building AR applications (Kato & Billinghurst, 1999).

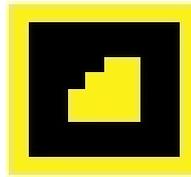


Figure 2. An example of a fiducial marker for vision based tracking.

By tracking and identifying markers placed in the environment, the algorithm calculates the position of the camera in relation to the markers, and hence the virtual information can be placed in the display relative to the marker position. Figure 3 illustrates how a marker (seen on the user's finger) is used to place the virtual 3D object in the user's field of view.

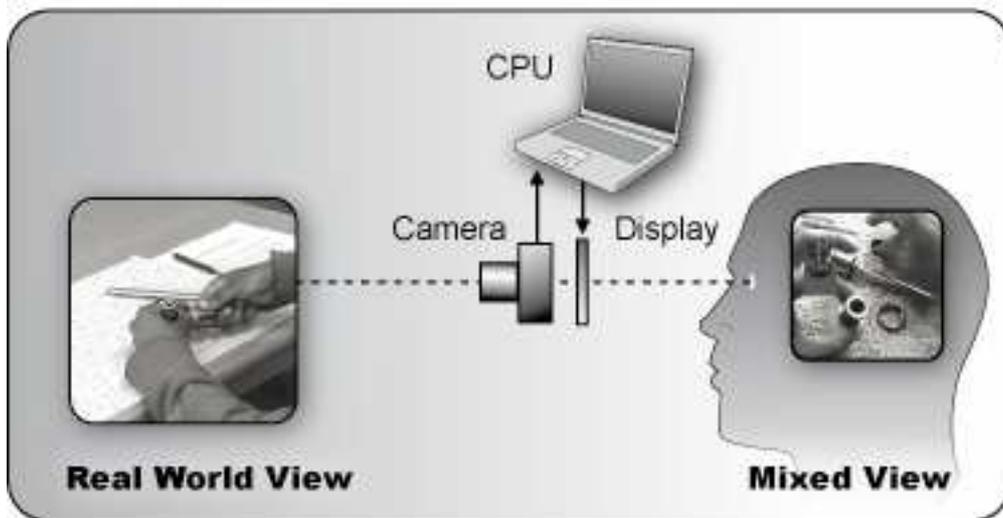


Figure 3. A schematic view of an AR system using marker tracking. Note the marker on the user's index finger.

One problem that can occur when using AR systems with marker based tracking is that when two markers are in the user's field of view, information is displayed at different locations at the same time. If the user only needs to see the information on one marker at a time, the system should be able to recognise this and switch off the

“unnecessary” information until it becomes significant for the user to see it. (In the case of the image above, the information related to the top marker is not relevant for the current task). This problem could perhaps be solved in a variety of ways, through hard coding, tracking technology and different input and interaction modalities. An alternative is to solve the problem by adding some sense of “user awareness” to the system, i.e. letting the AR system know what area and information the user is interested in at the moment. One possible way to do this is by making the AR system aware of the user’s visual interest with the aid of gaze tracking devices.

Adding user awareness to the system is an even more interesting approach in the promising emerging technique called markerless or feature based tracking. Markerless tracking will create opportunities to build more flexible systems where the user can view the world freely and still see virtual elements without being constrained to predefined areas in the world (Klein & Murray, 2007).

4. Gaze-based Interaction

For over twenty years, eye gaze interaction has been used effectively for eye typing, i.e. interacting with word processors through eye gaze instead of the traditional keyboard and mouse input (Majaranta & Rähkä, 2002). For a gaze-based interaction system to be useful, the gaze detection process should be implemented in a way that does not interfere with the user’s behaviour (Ohno & Mukawa, 2004). A gaze tracker should be able to work in different display configurations and in varying lighting conditions and, therefore, a gaze tracker is often a complex system. Many technologies are involved in a gaze tracker: displays, optic devices, image sensors and real-time signal processing. Each technology, with its characteristics, can affect the use of the tracker. For example, the gaze tracker’s accuracy and especially the system’s time delay are important in real-time interaction applications.

4.1 Eye Gaze and Usability

Usability testing methods presently involve a lot of trial and error when developing new interfaces. By using eye tracking, much time can be saved by directly giving answers to questions of layout, control functionality and visibility of different objects (Karn, Ellis & Juliano, 1999). Eye gaze tracking has been used to evaluate system usability as well as in psychological studies of human visual behaviour for a long time

(see for example Hyrskykari, 1997; Oviatt & Cohen, 2000; Maglio & Campbell, 2003; Qvarfordt, 2004).

A number of studies have shown that eye gaze patterns differ between expert and novice users of a system or method (Law, Atkins, Kirkpatrick, & Lomax, 2004). This implies that it is possible to use gaze patterns to see how a user interacts with a system, and the changes in gaze patterns may also reflect how the user's skills improve during training. This would make it possible to use changes in gaze patterns to evaluate the effectiveness of a training method or the design of an interface. Gaze recognition and logging can be a useful method for evaluating and developing AR applications for training and technical maintenance.

As noted previously, gaze behaviour can provide a system with implicit input, and this way of tracking the intentions of a user has been used successfully in multi-modal applications (Qvarfordt, 2004; Maglio & Campbell, 2003). In these studies, eye gaze direction was used to solve ambiguity problems in, for example, speech interaction: By registering on which visual area of the interface the user focuses, speech input can be interpreted faster and with more accuracy. In the example of a map, the visual area put some constraints on the possible input, which in this case would be names on the map. This is a way of adding an environmental context, the area of visual interest, to improve speech recognition (Shell, Selker & Vertegaal, 2003).

4.2 Problems with Gaze-controlled Interaction

Several studies discuss the potential problems of eye-gaze-based interaction (Vertegaal, 2002; Ohno, 1998; Zhai, Morimoto & Ihde, 1999; Zhai, 2003). One example of these problems is the "Midas touch" problem, which occurs when the user looks at an object without meaning to choose it but activates it anyway just by the first glance (Jakob, 1991; Hyrskykari, 1997; Ohno, 1998). Another problem occurs when using dwell time to determine whether a user has "clicked" on an object or not, since it is hard to define what the exact time should be before the object is activated (Majaranta, Aula, & R ih a, 2004). If the time chosen is too short, the "Midas touch" problem occurs; if the time is too long, the user may become annoyed. In both cases, the user will probably consider the interface to be quite user-unfriendly and will be reluctant to use it again (Ohno, 1998; Jakob, 1991). If dwell time is used as a means of activating objects on the screen or in typing with eye gaze, it is important that the feedback to the user is sharp and unambiguous (Majaranta, Aula, & R ih a, 2004).

An alternative way to interact visually with the system is to use eye gestures such as winks, but this is very difficult since the system has to be able to differentiate between

voluntary eye gestures and natural, involuntary eye gestures (Ohno, 1998). An additional important limitation of using gaze movements to interact with a system is that gaze will probably indicate visual interest, but not necessarily the cognitive interest of the user. It is one thing to determine if the user has had the information in the visual field of attention but a completely different issue to determine whether that information actually has been acknowledged and understood (Vertegaal, 2002; Bates & Istance, 2002). This could possibly be solved to some extent by requesting feedback from the user.

5. A Gaze-controlled Augmented Reality System

As stated above, gaze control as a method for both usability studies as well as interaction is not a new research domain. However, gaze-controlled interaction is new to the Mixed Reality domain. The main focus of implementing gaze control in a head-mounted mixed reality display is to make the interaction between the system and the user easier and more efficient. Integrating gaze detection into the AR system could be a way to predict the user's intention and to anticipate what actions will be requested of the system. The use of gaze control in the AR system will also be useful for developing the system's interaction methods and designing the displays. The user's eye movements can be used to interpret what the problem might be when the user does not handle the interface in the way it was intended (Karn, Ellis & Juliano, 1999). By analyzing the gaze patterns, one might be able to see if the user has observed objects or if the gaze is distracted by other objects.

In interactive AR applications that require responses from the user, there must be an efficient and non-interruptive way to deliver responses to the AR system. Gaze control in the AR system could make interaction between the system and the user easier and more efficient.

For an AR system using gaze-based interaction to be useful, the gaze detection process should be implemented in a way that does not interfere with the user's normal behavior (Oviatt & Cohen, 2000). A gaze tracker for AR applications should be able to integrate with micro-displays and must function in varying conditions of illumination. The following sections present a helmet-mounted AR system with an integrated gaze tracker, which can be used both to monitor the user's gaze behaviour as well as for interaction. The system has been tested to reach stage 2 in the testing process

described by Sommerville (2004). The first stage is component testing, and the third stage of the process is the acceptance testing stage, where a more elaborate usability study would take place.

5.1 HMD and Integrated Gaze Tracker

We have developed a head-mounted video see-through AR system with an integrated gaze tracker (see Figures 4 and 5). The integrated head-mounted display, black/white gaze tracker camera (640 × 480 pixels resolution) and scene camera is an in-house construction, and the different components used are illustrated in Figure 4.

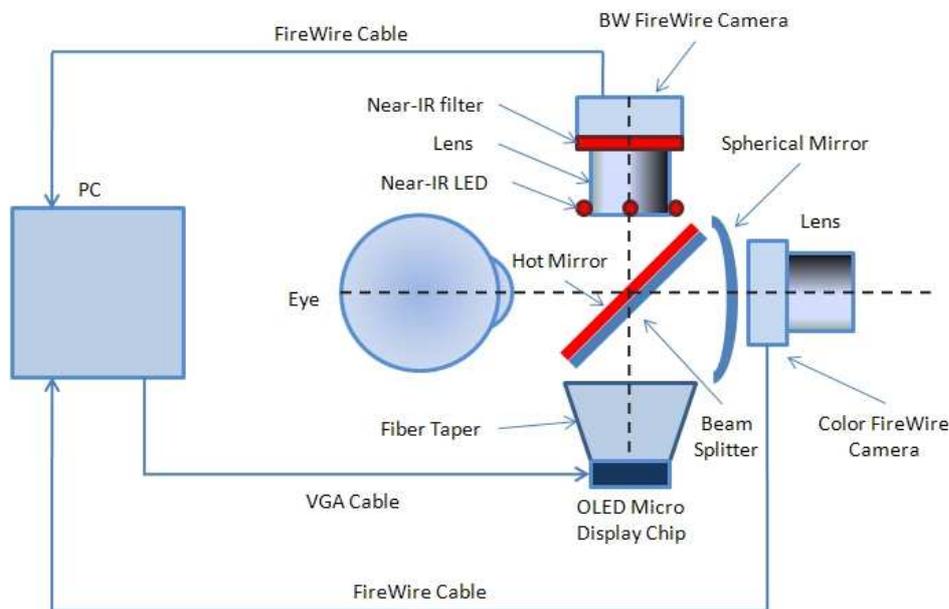


Figure 4. A schematic view of the gaze-controlled AR system.

The micro-displays have a resolution of 800 × 600 pixels and a field of view of 37 × 28 degrees. The gaze tracker camera and the micro-display are integrated and have co-aligned optic axes to facilitate future studies of vergence-movement-controlled systems (Gustafsson, Carleberg, Sohlberg, Persson, & Aldrin, 2005b).

The system utilizes the dark pupil tracking principle, which means that the NIR (near infrared) illumination is placed off-axis in relation to the gaze tracker (BW) camera (see Figure 5). Gaze tracking techniques generally use an illumination source, which may be placed either on- or off-axis with the gaze tracker camera. When the source is on-axis, the camera sees the pupil illuminated similar to the red-eye effect in flash photography. Thus on-axis tracking is also referred to as bright-pupil tracking. Similarly, off-axis tracking is referred to as dark-pupil tracking (Duchowski, 2007; Weigle & Banks, 2008; Young & Sheena 1975). The NIR illumination source is fully integrated

into the system and is not a separate device. The gaze tracker camera detects the pupil and its reflections by filtering and thresholding the image information. The position of the pupil and the positions of the reflections on the cornea caused by the NIR illumination are calculated. These positions are used as input parameters to a rather straightforward geometric algorithm that computes the gaze direction. Four illumination sources are used; however, only one reflection is needed to calculate gaze. The system can choose between the four different reflections, which increases the robustness of the system. Gaze is calculated using information on display geometry, camera placement and lens characteristics as well as initial assumptions on eye geometry. Interactive tuning/calibration is necessary in this prototype system, but this is normally a fast procedure (taking a few seconds) and could be automated.

5.2 Augmented Reality Software

The AR system described here uses hybrid tracking technology, basically a marker detection system based on ARToolKit (HITLAB webpage), ARToolKit Plus (Schmalstieg, 2005) and ARTag (Fiala, 2005) but with the addition of a 3DOF (3-degrees-of-freedom) inertial tracker (isense.com and xsens.com).

The software allows applications to be built and defined in a scenario file in XML syntax. For gaze controlled interaction, the application designer can define the layout of the gaze control dialog areas as well as gaze action specifications. With this tool, the application developer can experiment with, compare and verify the functionality of different gaze-controlled interaction schemes.

In the system that was developed, eye gaze interaction can be restricted both temporally and spatially – only certain parts of the display will have a function, and only when there is a need for gaze interaction. These interaction areas are defined in the application scenario XML file, which is also where eye gaze dwell times and command actions are configured. The gaze dialog area positions can either be fixed or dynamic, i.e. relative to a detected marker position, which allows flexible design of the application. The areas in which the gaze interaction is active are represented by transparent images, as can be seen in Figures 5 and 6. Transparency, color, appearance and placement of the interaction areas are defined in XML syntax in the scenario file.

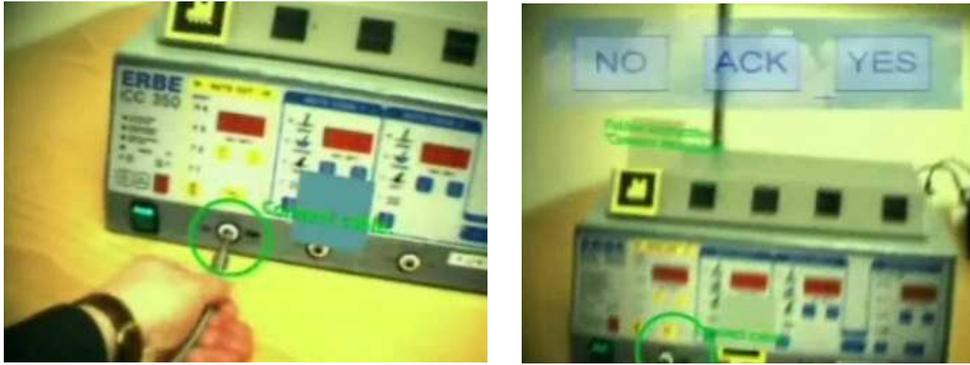


Figure 5. Gaze interaction areas can be concealed until the user needs them, thereby avoiding unnecessary clutter in the user's field of view.

The interactive area in the display can, as noted, be defined in XML syntax. The implementation allows several different ways of interacting with the system through gaze control. As an example, in one implementation, the central and peripheral areas of the display are defined as active and non-active interaction areas, respectively. This means that when the user's gaze is in the central part of the display, no interaction areas are displayed, as seen in the left-hand image in Figure 5. When the user wants to interact with the system, she/he gazes at the peripheral part of the display, which activates the interaction area in the display, as seen in the right image in the same figure. The software solution is described in detail in Gustafsson, Carleberg, Sohlberg, Persson and Aldrin (2005b).

6. Preliminary Test of the System

The gaze-controlled AR system was tested in a laboratory setting using two different instructional applications. Both applications were designed and evaluated previously, but without gaze control as an interaction device (see for instance Nilsson & Johansson, 2006; Nilsson & Johansson, 2008). Both applications were developed in cooperation with the target user group, which was medical staff at a hospital, and the task for the user was to complete a set of instructions. Three different settings of the gaze interaction dialogue were tested: a static interaction area in the upper part of the display, a static interaction area in the lower part of the display and, finally, a dynamic interaction area in which the interaction area appears relative to the placement of a marker. Limiting the interaction area to the upper part of the user's field of view solves part of the "Midas touch" problem, since users prefer not to use this part of the display in natural interaction. For example, Mon-Williams, Plooy, Burgess-Limerick, and Wann,

(1998) discuss problems with the visual angle of displays in VR). However, since looking at this area requires an uncomfortable gaze angle, users may not appreciate this variant, even though it solves the Midas touch problem. Placing the interaction dialogue in the lower part of the display instead is more gaze-friendly, but the problem of accidentally activating the gaze interaction dialogue is more prominent. To address this issue, the area that activates the dialogue was reduced in size to only two smaller boxes, one in each corner, as can be seen in Figure 7. When the user looks at one of the areas, this is indicated by a change in appearance (colour and image), so that the user receives feedback, acknowledging that the system “knows” that s/he is looking at the area.

The dynamic layout of the interaction dialog allows the application designer to link the gaze interaction directly to a marker, for instance, thus allowing interaction in the central field of view of the user when necessary. In the applications described, the point of gaze fixation is only visible when it is positioned in the interaction area in the upper/lower part of the display. However, it is also possible to show the gaze fixation point at all times if desirable. Interaction feedback is given to the user by changing the image and colour of the button (see Figure 6). This change indicates that the “press” has been acknowledged by the AR system.

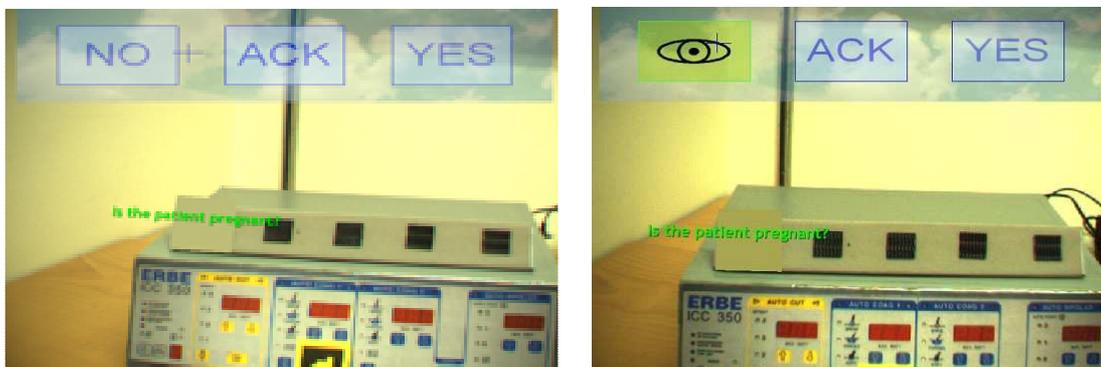


Figure 6. An example of the gaze interaction dialog. The user can respond to the question asked by looking at the different regions/“buttons” in the upper part of the field of view.

6.1 The First Application Example

This task has previously been used in a user study investigating the usability of AR instructions in the medical domain (Nilsson & Johansson, 2006, 2008). The main goal of the instructions is to activate an electro-surgical generator (ESG) and prepare it for operation. An ESG, often referred to as an “electrical knife”, is used for deep heating of tissue and is employed during surgery to minimise the amount of bleeding in the patient.

The instructions were given as statements and questions that had to be confirmed or denied via the input device, in this case the gaze interaction dialog, where the user can choose to look at “yes”, “no” and “ack” (short for “acknowledged”) (see Figure 6). The dwell time used in the test series was set to 1 second.

6.2 The Second Application

The task in the second application example has also been used previously in user studies of AR instructions in the medical domain (Nilsson & Johansson, 2007, 2008). In this application, the goal was to follow instructions on how to put a trocar together. A trocar is used as a “gateway” into a patient during minimal invasive surgeries. The trocar is relatively small and consists of seven separate parts, which have to be assembled correctly for it to function properly as a lock preventing blood and gas from leaking out of the patient’s body. The trocar was too small to have several different markers attached to each part. Markers attached to the object (such as the ones in study 1) would also not be realistic considering the type of object and its usage – it needs to be kept sterile and clean of other materials. Instead, the marker was mounted on a small ring with adjustable size, which the participants wore on their index finger (see Figure 7).



Figure 7. The gaze-controlled AR system in the second application. The images to the right show what the user sees. The “no” represents repeating an instruction (going back in the sequence), and the “yes” means going to the next step. Note that, in this setting, the interactive area is placed in the lower part of the display rather than the upper part as in the examples above.

The AR system used in this application was previously equipped with speech recognition and was controlled by verbal interaction from the user. The application,

which was originally designed for voice interaction, was redesigned and used in the gaze-controlled AR system. In contrast to the other application, in this task the users did not have to answer any questions but only needed to follow the instructions given, and the only input to the system was to acknowledge when an instruction was completed. After seeing the visual instruction and completing it, the user used gaze to let the system know if he/she wanted to repeat the instruction or go to the next step (see Figure 7).

The gaze-controlled AR application was tested in a laboratory setting and was not a full user study, but rather a functionality test. Trials that were performed with the system indicated that the system functions as intended.

6.3 Results of the Preliminary Tests

During the trials, it was found that the users tended to turn and tilt their head so that the focus of attention was always in the centre of the field of view. Looking at things in the upper section of the display was therefore a conscious effort and probably not part of casual gazing. Users reported that it was strenuous to interact with this gaze angle and preferred the lower static interaction dialogue. There were, however, some problematic issues with the lower interaction dialogue as well: When the users tried to activate the dialogue, they tended to tilt their head, thus often losing camera contact with the marker, which in turn led to the loss of the visual instructions. In conclusion, the static dialogues are not ideal since they are not adapted to normal human behaviour: When humans want to focus their gaze on an object, virtual or real, they tend to place it in the central field of view. This is of course not possible when the instruction dialogue is static.

The dynamic dialogue does not have the problem of the users tilting and moving their heads. However, one important problem was found. Although the interaction dialogue was only visible when the system required an input from the user, it still sometimes covered too much of the user's field of view. This caused the participants to experience it as being cluttered. This could be addressed with a redesign of the interaction dialogue in future development of the system.

In general, the participants reacted positively to the concept of gaze control in these types of applications, but the system was experienced as clumsy and not entirely stable, since they sometimes lost the virtual information when the marker was not detected by the camera. The latter problem can be addressed by further refining the AR/MR system and the software. The clumsiness of the system is harder to address in the technical solution presented here. Video-see-through AR and gaze control requires

cameras (two at a minimum), and these are too heavy to be comfortably placed on a regular head-mounted setup – the helmet helps to balance the weight of the system. For gaze-controlled video see-through AR, a helmet-mounted solution is currently the best option.

7. Discussion and Future Research

As mentioned previously, this paper presents a system test, which means that the focus of the study was to test the functionality of the gaze-controlled application, as opposed to a more elaborate end user study, where more aspects than the technical functionality are of interest. Nevertheless, the test resulted in some preliminary insights into the user aspects of the system. The gaze interaction in the applications tested was basically the same as with ordinary keyboard buttons. The experience so far is that gaze controlled interaction is as fast and distinct as pressing an ordinary keyboard button. This is in accordance with earlier research and the results of Ware and Mikaelian (1986), who showed that gaze interaction may even be faster because the time it takes to shift the position of the cursor manually causes a delay. Nevertheless, there are still many areas in which the system can be improved. The main factor is the size of the system – the cameras and displays still require a helmet-mounted set-up in order for calibration and weight distribution to work properly. The system would also require some more extensive user studies to ensure the effectiveness of the gaze control, and also to evaluate the system not only in technical terms but also more importantly in terms of end user satisfaction.

The use of dwell time as an interaction method was a choice made during the development process; to decrease the risk of the Midas touch phenomenon, the system also gives the user visual indicators of the gaze response. That is, when the user has dwelled on a gaze button area for about a half second, the area changes its appearance to make the user aware of the interaction activity. This is one way to give clear feedback to the user. The dwell times used to determine selection in the interaction have not yet been fully investigated but rely on previous studies in the domain of eye gaze interaction.

There may also be alternatives to using dwell time (as in the described system) and eye gestures such as winks, gaze gestures (Drewes & Schmidt, 2007) and “eye graffiti” (Porta & Turina, 2008) that have not been investigated for use in this application as of

today. For instance, eye gestures other than winks could be used. Using eye gaze patterns that can easily be identified as voluntary input instead of dwell time or winks could improve the usability. An example of such voluntary eye gestures could be moving the gaze quickly in one direction and then back again. This is similar to the gestures used with other pointing devices, such as the pointer gestures used for typing on small screens such as those of mobile phones.

Gaze interaction allows the user to work freely with her/his hands while stepping through an instruction hierarchy. This freedom of movement is of value in many situations, not only in the application described above but also in other applications involving maintenance and repair tasks. The conclusion therefore is that the system can be used as an alternative to traditional manual interaction. This is of particular interest for applications where manual input or speech input is not appropriate or possible. The experiences from the limited test runs are important for the further development of the system and have clearly indicated that the system is functional. Future tests, including a larger user group, will investigate the robustness of the system as well as give more insight into the speed and accuracy in other applications than the hierarchical instructions used here.

The gaze recognition in the system is not restricted to use for direct interaction but can also be used for indirect communication with the AR system. Gaze recognition can add a “user awareness” dimension to the system, which can monitor the user’s visual interest and act upon this. Gaze awareness can also allow the system to infer (via the user’s gaze direction) when and if the user wants to interact with the system. If the user has two or more markers in the field of view, the gaze direction can be used to indicate which marker’s virtual information should be displayed and is thus a relatively easy way to de-clutter the user’s field of view.

Combining the concepts from AR with gaze recognition and input enables quick and easy interaction in an AR system that allows for natural human communication, such as communicating intention by the use of eye gaze. Cognitive interest may not always be the same as visual interest, but, in many cases, visual interest can be an indicator of what the user is focusing on both visually and cognitively and can therefore allow the system to respond to this, for instance by presenting requested virtual information.

The proposed AR system with added gaze awareness and gaze-controlled interaction will be further tested in a user application. Another aspect of the gaze-aware AR system is its ability to log and analyze the gaze patterns of AR users, possibly allowing further usability studies and evaluations of the AR system. Gaze control in the AR-

system may also be a useful tool for developing the system's interaction methods and the designing the displays.

8. Acknowledgments

This research was a collaboration project between the department of Computer and Information Science (IDA) at Linköping University and the Swedish Defence Research Agency, funded by the Swedish Defence Materiel Administration (FMV).

9. References

- Azuma, R. (1997). A survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355 – 385.
- Azuma, R., Bailot, Y., Behringer, R., Feiner, S., Simon, J. & MacIntyre, B. (2001). Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 21(6), 34-47.
- Bates, R. & Istance, H. (2002). Zooming interfaces! Enhancing the performance of eye controlled pointing devices. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies* (pp. 119-126). New York: ACM Press.
- Billinghurst, M., Kato, H. & Poupyrev, I. (2001). The MagicBook: Moving Seamlessly between Reality and Virtuality. *IEEE Computer Graphics and applications*, 21(3), 2-4.
- Drewes, H. & Schmidt, A. (2007, September). Interacting with the Computer using Gaze Gestures. Presented at *Interact 2007*, Rio De Janeiro, Brazil.
- Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and practice*. London: Springer.
- Fiala, M. (2005, October). ARTAG Rev2 Fiducial Marker System: Vision based Tracking for AR. Presented at the *Workshop of Industrial Augmented Reality*, Vienna, Austria.
- Gandy, M., MacIntyre, P., Presti, P., Dpw, J. Bolter, B. Yarbrough, & O'Rear, N. (2005). AR Karaoke: Acting in your favourite scenes. In *Proceedings of the fourth IEEE and ACM International conference on Mixed and Augmented Reality, ISMAR* (pp. 114-117). California: IEEE Computer Society.

- Gustafsson, T. & Carleberg, C. (2003). *Mixed Reality for Technical Support*. Technical report FOI--0857--SE, Swedish Defence Research Agency.
- Gustafsson, T., Carleberg, P., Nilsson, S., Svensson, P., Sivertun, Å. & LeDuc, M. (2004). *Mixed Reality for technical support*. Technical report ISSN 1650-1942, Swedish Defence Research Agency.
- Gustafsson, T., Carleberg, P., Svensson, P., Nilsson, S., & Sivertun, Å. (2005a). *Mixed Reality Systems for Technical Maintenance and Gaze-controlled interaction*. Technical report ISSN 1650-1942, Swedish Defence Research Agency.
- Gustafsson, T., Carleberg, P., Sohlberg, P., Persson, B. & Aldrin, C. (2005b). *Interactive Method of Presenting Information in an Image*. No WO/2005/124429, patent granted 2005-05-10.
- Haller, M., Billinghamurst, M., & Thomas, B. (Eds) (2007). *Emerging Technologies of Augmented Reality: Interfaces and Design*. London, UK: Idea Group Publishing.
- Henrysson, A., Ollila, M. & Billinghamurst, M. (2007). Mobile Phone Based Augmented Reality. In Haller, M., Billinghamurst, M. & Thomas, B. (Eds.), *Emerging technologies of Augmented Reality. Interfaces and design* (pp. 90-109). London: Idea Group Publishing.
- HITLAB, <http://www.hitl.washington.edu/artoolkit/> (as of 2008-06-10).
- Hyrskykari, A., Majaranta, P. & R  ih  , K.J. (2005, July). From Gaze Control to Attentive Interfaces. Presented at the *3rd International Conference on Universal Access in Human-Computer Interaction*, Las Vegas, NV, USA.
- Hyrskykari, A. (1997). *Gaze Control as an Input Device*. Report B-1997-4. University of Tampere, Department of Computer Science. Retrieved on-line August, 15, 2009 from: <http://www.cs.uta.fi/hci/gaze/publications.php>
- Jakob, R.K. (1991). The use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. *ACM Transactions on Information Systems*, 9(3), 152-169.
- Karn, K.S., Ellis, S. & Juliano, C. (1999). The Hunt for Usability: Tracking Eye Movements. In *Proceedings of CHI'99* (pp. 173-173). New York: ACM Press.
- Kato, H. & Billinghamurst, M. (1999, October). Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. Presented at the *2nd International Workshop on Augmented Reality, IWAR 99*, San Francisco, USA.
- Kiyokawa, K. (2007). An Introduction to Head Mounted Displays for Augmented Reality. In Haller, M., Billinghamurst, M. & Thomas, B. (Eds.). *Emerging Technologies*

- of Augmented Reality: Interfaces and Design* (pp. 43-63). Idea Group Publishing, London, UK.
- Klein, G. & Murray D. (2007). Parallel Tracking and Mapping for Small AR Workspaces. In *Proceedings of the sixth IEEE and ACM International conference on Mixed and Augmented Reality* (pp. 225--234). Washington DC: IEEE Computer Society.
- Law, B., Atkins, M. S., Kirkpatrick, A. E. & Lomax, A. J. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the Eye Tracking Research and Application Symposium* (pp. 41-48). New York: ACM Press.
- Maglio, P. & Campbell, C. (2003). Attentive Agents. *Communications of the ACM*, 46(3), 47-51.
- Majaranta, P. & R ih a, K-J. (2002). Twenty years of eye typing. In *Proceedings of Eye Tracking Research and Applications, ETRA'02* (pp. 15-22). New York: ACM Press.
- Majaranta, P., Aula, A. & R ih a, K-J. (2004). Effects of Feedback on Eye Typing with Short Dwell Time. In *Proceedings of Eye Tracking Research and Applications, ETRA'04* (pp.139-146). New York: ACM Press.
- Milgram, P. & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems, E77-D(12)*, 1321-1329.
- Mon-Williams, M., Plooy, A., Burgess-Limerick, R & Wann, J. (1998). Gaze angle: A possible mechanism of visual stress in virtual reality headsets. *Ergonomics*, 48(3), 280-285.
- Nilsson, S. & Johansson, B.J.E. (2006). A cognitive Systems Engineering Perspective on the Design of Mixed Reality Systems. In *Proceedings of the 13th European Conference on Cognitive Ergonomics* (pp. 154-161). New York: ACM
- Nilsson, S. & Johansson, B. (2007). Fun and usable: Augmented reality instructions in a hospital setting. In *Proceedings of the 19th Australasian Conference on Computer-Human interaction: Entertaining User interfaces* (pp. 123-130). New York: ACM Press.
- Nilsson, S. & Johansson, B.J.E. (2008). Acceptance of Augmented Reality Instructions in a Real Work Setting. In *Proceedings of Conference on Human Factors in Computing Systems* (pp 2025-2032). New York: ACM Press.
- Ohno, T. (1998). Features of Eye Gaze Interface for Selection Tasks. In *Proceedings of the 3rd Asia Pacific Computer-Human Interaction, APCHI'98* (pp.176-182). Washington, DC: IEEE Computer Society.

- Ohno, T. & Mukawa, N. (2004). A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction. In *Proceedings of Eye Tracking Research and Applications, ETRA'04* (pp. 115 – 122). New York: ACM Press.
- Oviatt, S. & Cohen, P. (2000). Multimodal Interfaces that Process What Comes Naturally. *Communications of the ACM*, 43(39), 45-53.
- Porta, M. & Turing, M. (2008, March). Eye-S: a Full-Screen Input Modality for Pure Eye-based Communication. Presented at *the 2008 Symposium on Eye Tracking Research and Applications*. Savannah, GA, USA.
- Qvarfordt, P. (2004). *Eyes on Multimodal Interaction*. Linköping Studies in Science and Technology, Dissertation no. 893. Linköping: Unitryck.
- Sareika, M. (2005). *Einsatz von Eye-Tracking zur Interaktion in Mixed Reality Umgebungen*. Diplomarbeit. Fraunhofer Institut Angewandte Informationstechnik.
- Schmalstieg, D. (2005, October). Rapid Prototyping of Augmented Reality Applications with the STUDIERSTUBE Framework. Presented at the *Workshop of Industrial Augmented Reality*, Vienna, Austria.
- Shell, J. S., Selker, T., & Vertegaal, R. (2003). Interacting with groups of computers. *Communications of the ACM*, 46(3),40-46.
- Sommerville, I. (2004). *Software Engineering*. Edinburgh Gate: Pearson Education.
- Stanney, K. (1995, March). Realizing the full potential of virtual reality: human factors issues that could stand in the way. Presented at the *Virtual Reality Annual International Symposium, VRAIS'95*, Research Triangle Park, NC, USA.
- XSens: <http://www.xsens.com/> (as of 2008-06-10)
- Vertegaal, R. (2002). Designing Attentive Interfaces. In *Proceedings of the 2002 Symposium on Eye Tracking Research and Applications, ETRA'02* (pp. 23-30). New York: ACM Press.
- Ware, C. & Mikaelian, H.H. (1986). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCHI/GI conference on Human Factors in computing systems and graphic interfaces* (pp. 183-188). New York: ACM Press.
- Weigle, C. & Banks, D.C. (2008, January). Analysis of eye-tracking experiments performed on a Tobii T60. Presented at the *Conference on Visualization and Data Analysis*, San José, California, USA
- Young, L., R. & Sheena, D. (1975) Survey of eye movement recording methods. *Behavior Research methods & Instrumentation*, 7(5), 397-429.
- Zhai, S. (2003). What's in the eyes for attentive input? *Communications of the ACM*, 46(3), 34 – 39.

Zhai, S., Morimoto, C., and Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In *Proceeding of SIGCHI Conference on Human Factors in Computing Systems* (pp. 246-253). New York: ACM Press.