A Holistic Approach to Design and Evaluation of Mixed Reality System

Susanna Nilsson, Björn Johansson and Arne Jönsson

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-52595>

N.B.: When citing this work, cite the original publication.

The original publication is available at www.springerlink.com:

Nilsson, S., Johansson, B., Jönsson, A., (2010), A Holistic Approach to Design and Evaluation of Mixed Reality System, *The Engineering of Mixed Reality Systems*.
https://doi.org/10.1007/978-1-84882-733-2_3

Original publication available at:

https://doi.org/10.1007/978-1-84882-733-2_3

Copyright: Springer

<http://www.springer.com/>

Chapter 1

A holistic approach to design and evaluation of Mixed Reality systems

Susanna Nilsson, Björn Johansson, Arne Jönsson

Department of Computer and Information Science, Linköping University, Sweden,
Saab Security, Sweden, and Santa Anna IT Research Institute AB

Abstract This chapter addresses issues related to usability and user experience of Mixed Reality (MR) systems based on a naturalistic iterative design approach to the development of MR applications. Design and evaluation of MR applications are still mostly based on methods used for development of more traditional desktop graphical user interfaces. MR systems are in many aspects very different from desktop computer applications so these traditional methods are not sufficient for MR applications. There is a need for new approaches to user centered design and development of MR systems. One such approach is based on the concepts of Cognitive Systems Engineering (CSE). In this chapter we show how this approach can be applied to the development of MR systems. Two case studies are described, where a holistic CSE approach to design, implementation and evaluation has been used. The results show that allowing real end users (field/domain experts) to interact in a close to naturalistic setting provides insights on how to design MR applications that are difficult to attain otherwise. We also show the importance of iterative design, again involving real end users.

Key words: mixed reality systems, augmented reality, user study, user evaluation

1.1 Introduction

Mixed Reality (MR) systems are still, 40 years after Ivan Sutherland's [34] first descriptions of a head mounted display, mainly used and studied in the research domain. There are no practical and easy-to-use commercial off-the-shelf MR applications, or widely used systems in the same way that many other computer based applications have become a natural part of everyday working life in the industrialized world. New applications are developed every year, but to a large extent they remain in the research community, seemingly far away from the potential end users.

There can of course be many reasons for this development, but one issue that contributes to this fact is that so few of the designs are actually based on explicit user needs and requirements. There are few examples of MR applications that have been developed solely as a way to solve a real end user problem. Instead many applications are developed mainly because the technical resources and skills to develop them exist, alongside researchers with creative ideas. This type of research is important, but the resulting applications and designs are then not necessarily related to real end users and their needs [11]. Furthermore, these systems are often at best tested on a number of more or less randomly chosen people and evaluated based on participant comments, statistical measurements, or technical findings [35, 7].

The methods used for user evaluation of MR applications are in general based on usability methods for graphical user interfaces, sometimes in combination with usability for Virtual Reality (VR) applications [36]. Currently most papers in the field of MR which include evaluation of some kind (not necessarily user evaluation) are quantitative, and of the ones referred to as user evaluations the vast majority make use of objective measurements such as error rate, task completion time etcetera [7] and the participants tend to be randomly chosen students. These studies stem from a more traditional view of cognition where the use of structured experiments and quantitative measurements is common. This approach, although the most commonly used, may not be the most appropriate. We do, however, not claim that studies on performance and economic relevance are unnecessary, but that such studies are not enough to understand how users experience and use MR-systems.

As noted by Livingston [21] and Nilsson and Johansson [25] there are problems addressing human factors in MR systems. MR systems and applications differ from standard desktop applications in many ways and these differences between MR systems and desktop computer display based systems create a need for a different approach to both development and evaluations of these systems. To understand the potential of MR systems in real world tasks the technology must be designed based on investigations in real world scenarios.

This chapter will describe and discuss two user applications where the studies to some extent represent examples of traditional qualitative user studies. The first study illustrates the approach applied to a single user application. This study has been published previously to varying extent [26, 27]. The second study illustrates the approach in a collaborative setting. This study has not been reported previously. The participants involved in both studies are real end users – in the first study medical staff at a hospital, and in the second study personnel from three civil service organisations – and the applications have been built after explicit requirements and involvement from the particular end user groups. The focus of the result analysis is not traditional quantitative values, but rather qualitative values such as user experience and acceptance [5].

1.2 Related work

In this section we describe properties of MR systems and important aspects on how to evaluate and design these systems.

1.2.1 Mixed Reality systems

The general aim of Mixed Reality systems is the merging of worlds by adding virtual information to the real world. The field of Mixed Reality is a relatively new field in terms of commercially and publicly available applications. As a research field, however, it has existed for a considerably longer time with applications in diverse domains, such as medicine, military applications, entertainment and infotainment, technical support and industrial applications, distance or remote operation and geographic applications [2, 1].

Milgram and Kishinos [22] virtual continuum is often used to describe the relation between augmented reality, virtual reality and the stages in between. Mixed Reality is the collective name for all the stages (see Figure 1.1).

Fig. 1.1 The Virtual Continuum (after Milgram and Kishino)

The systems used in the case studies in this chapter can be defined as MR systems as they hold the possibility of being completely immersive. We use head-mounted displays that can be used to display only virtual information, shielding the user off from the surrounding world. However, they can also be used as AR systems; the applications described in this chapter are not immersive, but rather augments the users normal sight with virtual elements. To be considered an AR system, the system has to fulfill three

criteria according to Azuma [1]: they all combine the real and the virtual, they are supposedly interactive in real time, and they are registered and aligned in 3D [2, 1]. These three criteria are fulfilled in the systems described in this chapter.

1.2.2 Usefulness

To successfully integrate new technologies into an organisation or workplace means that the system, once in place, is actually used by the people it is intended for. There are many instances where technology has been introduced in organisations but not been used for a number of different reasons. One major contributor to the lack of

usage is of course the usability of the product or system in itself. But another issue is how well the system operates together with the users in a social context – are the users interested and do they see the same potential in the system as the people (management) who decided to introduce it in the organisation? Davis [5] describes two important factors that influence the acceptance of new technology, or rather information systems, in organisations: the perceived usefulness of a system and the perceived ease of use. Both influence the attitude towards the system, and hence the user behaviour when interacting with it, as well as the actual use of the system. If the perceived usefulness of a system is considered high, the users can accept a system that is perceived as harder to use than if the system is not perceived as useful. For example, in one study conducted in the field of banking the perceived usefulness of the new technology was even more important than the perceived ease of use of the system, which illustrates the need to analyse usability from more than an ease-of-use perspective [19]. For an MR system this means that even though the system may be awkward or bulky, if the applications are good, i.e. useful enough, the users will accept and even appreciate it. Equally, if the MR system is not perceived useful, the MR system will not be used, even though it may be *easy* to use.

1.2.3 *Technology in context*

Introducing new technology in a specific domain affects not only the user, but also the entire context of the user, and most noticeably the task that the user performs. Regardless of whether the technology is only upgraded or if it is completely new to the user, the change as such will likely have an effect on the way the user performs his/her tasks. What the effects may be is difficult to predict as different users in different contexts behave differently. The behaviour of the user is not only related to the specific technology or system in question but also related to the organisational culture and context [37]. This implies that studying usefulness of technology in isolation from the natural context (as in many traditional, controlled usability studies) may not actually reveal how the technology will be used and accepted in reality. Understanding, and foreseeing the effects of change on both user, task and context requires knowledge about the system, but perhaps even more importantly, an understanding of the context, user and user needs [16].

Usability guidelines such as the ones presented by Nielsen [24], Shneiderman [32] or other researchers with a similar view of cognition and usability are often the main sources of inspiration for usability studies of MR and AR systems. Several examples of such studies are listed in the survey of the field conducted by Dünser *et. al.* [7]. The guidelines used in these studies are sensible and purposeful in many ways but they often fail to include the context of use, the surroundings and the effect the system or interface may have in this respect. Being contextually aware in designing an interface means having a good perception of not only *who* the user is but also *where* and *how* the system can and should affect the user in his/her tasks.

In many usability studies and methods the underlying assumption is that of a de-composed analysis where the human and the technical system are viewed as separate entities that interact with each other. This assumption can be accredited the traditional idea of the human mind as an information processing unit where input is processed internally followed by some kind of output [23]. In this view, cognition is something inherently human that goes on inside the human mind, more or less isolated. The main problem with these theories of how the human mind works is not that they necessarily are wrong, the problem is rather that they to a large extent are based on laboratory experiments investigating the internal structures of cognition, and not on actual studies of human cognition in an actual work context [23, 6].

Another issue that complicates development and evaluation of systems is what is sometimes referred to as the 'envisioned world problem' which means that even if a good understanding of a task exists, the new design, or tool, will change the task, making the first analysis invalid [17, 39].

As a response to this, a general and approach to human-machine interaction has been suggested by Hollnagel and Woods called *Cognitive Systems Engineering* (CSE) [16, 17]. The main idea in the CSE approach is the concept of cognitive systems, where the humans are a part of the system, and not only users of that system. The focus is not on the parts and how they are structured and put together, but rather the purpose, and the function of the parts in relation to the whole. This means that rather than isolating and studying specific aspects of a system by conducting laboratory studies, or experiments under controlled conditions, users and systems should be studied in their natural setting, doing what they normally do. For obvious reasons, it is not always possible to study users in their normal environment, especially when considering novel systems. In such cases, CSE advocates studies of use in simulated situations [16]. Thus, the task for a usability evaluator is not to analyse only details of the interface, but rather allowing the user to perform meaningful tasks in meaningful situations. This allows for a more correct analysis of the system as a whole. The focus of a usability study should be the user performance with the system rather than the interaction between the user and the system. Comprising the concepts derived from the CSE-perspective and Davis (presented above), the design of a system should be evaluated based both on how users actually perform with a specific artifact, but also how they experience that they can solve the task with or without the artifact under study.

CSE is thus in many respects a perspective comprising several theories and ideas rather than a single theory. A central tenant in the CSE perspective is that a cognitive system is a goal-oriented system able to adjust its behaviour according to experience [16]. Being a child of the age of expert systems, CSE refers to any system presenting this ability as a "cognitive system". Later, Hollnagel and Woods [16] introduced the notion of "joint cognitive system" pointing to a system comprised of a human and the technology that human uses to achieve a certain task in a certain context.

Several other cognitive theories such a situated cognition [33], distributed cognition [?] and activity theory [38, 8] advocates similar perspectives. What all of these approaches share is that they mostly apply qualitative methods such as observation

studies, ethnography, conversational analysis etc. Grounded in the perspective that knowledge can be gained in terms of patterns of behaviour that can be found in many domains, strictly quantitative studies are rarely performed. A basic construct in the CSE movement is the cyclic interaction between a cognitive system and its surroundings (see Figure 1.2). Each action performed is executed in order to fulfill a purpose, although not always founded on an ideal or rational decision. Instead, the ability to control a situation is largely founded on the competence of the cognitive system (the performance repertoire) and the available information about what is happening and the time it takes to process it.

Fig. 1.2 The basic cyclical model as described in CSE.

From the basic cyclical model, which essentially is a feedback control loop operating in an open environment, we can derive that CSE assumes that a system has the ability to shape its future environment, but also that external influence will shape the outcome of actions, as well as the view of what is happening. The cognitive system must thus anticipate the outcomes of actions and interpret the actual outcome of actions and adjust in accordance to achieve its goal(s). Control from this perspective is the act of balancing feed forward with feed back. A system that is forced into reactive feed back-driven control mode due to lack

of time or an adequate model of the world is bound to produce short-sighted action and is likely to ultimately lose control. A purely feed forward driven system would conversely get into trouble since it has to rely solely on world models, and these are never complete. What a designer can do is create tools that support the human agent in such a way that it is easier to understand what is happening in the world or amplify the human ability to take action, creating a joint cognitive system that can cope with its tasks in a dynamic environment. This is particularly important when studying MR systems as the outspoken purpose of MR is to manipulate perceptual and cognitive abilities of a human by merging virtual elements into the perceived real world. From the very same point of view a number of risks with MR can be identified; a poorly designed MR system will greatly affect the joint human-MR system's performance.

Most controlled experiments are founded on the assumption of linear causality, i.e. that external variables are controlled and that action A always leads to outcome B. In most real-world or simulated environments with some level of realism and dynamics, this is not the case. Depending on circumstances and previous actions, the very same action may yield completely different outcomes. Taking this perspective, *process* becomes the focus of study, suggesting that it is more important to produce accurate descriptions than using pre-defined measures. The pre-defined measures

demand that a number of assumptions about the world under study stays intact, assumptions that rarely can be made when introducing new tools such as MR systems in a real working environment.

Although some social science researchers [20] perceive qualitative and quantitative approaches as incompatible, others [28, 30] believe that the skilled researcher can successfully combine approaches. The argument usually becomes muddled because one party argues from the underlying philosophical nature of each paradigm, and the other focuses on the apparent compatibility of the research methods, enjoying the rewards of both numbers and words. Because the positivist and the interpretivist paradigms rest on different assumptions about the nature of the world, they require different instruments and procedures to find the type of data desired. This does not mean, however, that the positivist never uses interviews nor that the interpretivist never uses a survey. They may, but such methods are supplementary, not dominant. Different approaches allow us to know and understand different things about the world. Nonetheless, people tend to adhere to the methodology that is most in accordance with their socialized worldview [12, p. 9].

1.3 User involvement in the development process

A long term goal of MR research is for MR systems to become fully usable and user-friendly, but there are problems addressing human factors in MR systems [21, 27]. As noted, research in MR including user studies usually involves mainly quantifiable measures of performance (e.g. task completion time and error rate) and rarely focuses on more qualitative performance measures [7]. Of course there are exceptions such as described by Billingham and Kato [3] where user performance in collaborative AR applications is evaluated by not only quantitative measures but also with gesture analysis and language use. Another example is the method of domain analysis and task focus with user profiles presented by Livingston [21]. However these approaches are few and far between. For most instances the dominating method used is quantitative.

To investigate user acceptance and attitude towards MR technology, this chapter describes two case studies, one single user application and one collaborative multi-user application. The MR applications were developed iteratively with participation from the end user group throughout the process of planning and development. In the first study the participants received instructions on how to assemble a relatively small surgical instrument, a trocar, and the second study focused on a collaborative task, where participants from three different organisations collaborated in a command and control forest fire situation. While the first case study is an example of a fairly simple individual instructional task, the second case study is an example of a dynamic collaborative task where the MR system was used as a tool for establishing common ground between three participating organisations.

1.3.1 The method used in the case studies

Acknowledging the 'envisioned world' problem, we have adapted an iterative design approach where realistic exercises are combined with focus groups in an effort to catch both user behaviour and opinions. The studies have a qualitative approach where the aim is to iteratively design a MR system to better fit the needs of a specific application and user group.

As noted, the CSE approach to studying human computer interaction advocates a natural setting, encouraging natural interaction and behaviour of the users. A fully natural setting is not always possible, but the experiments must be conducted in the most likely future environment [11, 16]. A natural setting makes it rather meaningless to use metrics such as task completion time, c.f. [14] as such studies require a repeatable setting. In a natural setting, as opposed to a repeatable one, unforeseen consequences are inevitable and also desirable. Even though time was recorded through the observations in the first case study presented in this chapter, time was not used as a measurement, as time is not a critical measure of success in this setting. In the second setting task completion could have been used as a measurement but the setup of the task and study emphasized collaboration and communication rather than task performance. The main focus was the end users' experience of the system's usefulness rather than creating repeatable settings for experimental measurements. Experimental studies often fail to capture how users actually perceive the use of a system. This is not because it is impossible to capture the subjective experience of use, but rather because advocates of strict experimental methods often prefer to collect numerical data such as response times, error rates, etcetera. Such data is useful but it is also important to consider qualitative aspects of user experience.

1.3.2 The design process used in the case studies

The method used to develop the applications included a pre-design phase where field experts took part of a brainstorming session to establish the initial parameters of the MR system. These brainstorming sessions were used to define the components of the software interface, such as what type of symbols to use, and what type of information is important and relevant in the different tasks of the applications. In the single user application this was for instance the specifics of the instructions designed and in the multi user application the specifics of the information needed to create common ground between the three participating organisations. Based on an analysis of the brainstorming session a first design of the MR system as well as of the user task was implemented. The multi user application was designed to support cooperation as advocated by Billinghurst and Kato [3] and thus emphasized the need for actors to see each other. Therefore, we first used hand-held devices that are easier to remove from the eyes than head mounted displays, see Figure 1.3.

The first design and user task was then evaluated by the field experts in workshops where they could test the system and give feedback on the task as well as

on the physical design and interaction aspects of the system. In the single user application the first system prototype needed improvements on the animations and instructions used, but the hardware solutions met the needs of the field experts. The evaluation of the multi user application however illustrated several problematic issues regarding physical design as well as software related issues, such as the use of the hand-held display which turned out to be a hindrance for natural interaction rather than an aid.

Fig. 1.3 Digital pointing using an interaction device and hand held displays.

For the multi user application we further noticed that when a user points at things in the map, the hand is occluded by the digital image of the map in the display. Thus, hand pointing on the digital map was not possible due to the technical solution. This problem was solved by using an interaction device to point digitally (see Figure 1.3). The virtual elements (the map and symbols) and interaction device also had several improvement possibilities.

Based on this first prototype evaluation another iteration of design and development took place where the MR system went through considerable modifications. Modifications and im-

provements were also made on the user task. As mentioned the single user application and task only had minor changes made, while the multi user application went through considerable changes. The hand held displays were replaced with head mounted displays, the interaction device was transformed and the software was upgraded considerably to allow more natural gestures (hand pointing on the digital map). Besides these MR system related issues several changes were also made to the user task to ensure realism in the setting of the evaluation.

After the changes to the MR system and user task were made, another participant workshop was held to evaluate the new system design and the modified user task and scenario. This workshop allowed the field experts to comment and discuss the updated versions of the applications and resulted in another iteration of minor changes before the applications were considered final and ready for the end user studies. The final applications are described in detail in the following case study descriptions.

1.4 The First Case Study – an Instructional Task

The public health care domain has many challenges and among them is, as in any domain, the need for efficiency and making the most of available resources. One part of the regular activities at a hospital is the introduction and training of new staff.

Even though all new employees may be well educated and professionally trained, there are always differences in tools and techniques used – coming to a new work place means learning the equipment and methods used in that particular place. In discussions following a previous study [25] one particular task came up as something where MR technology might be an asset in terms of a training and teaching tool. Today when new staff (physicians or nurses) arrive it is often up to the more experienced nurses to give the new person an introduction and training to tools and equipment used. One such tool is the trocar (see Figure 1.4), which is a standard tool used during minimal invasive surgeries. There are several different types and models of trocars and the correct assembly of this tool is important in order for it to work properly. This was pointed out as one task that the experienced staff would appreciate not having to go through in detail for every new staff member. As a result MR instructions were developed and evaluated as described in this section.

1.4.1 Equipment used in the study

The MR system included a Sony Glasstron head mounted display and an off the shelf headset with earphones and a microphone. The MR system runs on a laptop with a 2.00 GHz Intel®Core™2 CPU, 2GB RAM and a NVIDIA GeForce 7900 graphics card. The MR system uses a hybrid tracking technology based on marker tracking; ARToolKit, (available for download at [15]), ARToolKit Plus [31] and ARTag [9]). The marker used can be seen in Figure 1.4. The software includes an integrated set of software tools such as software for camera image capture, fiducial marker detection, computer graphics software and also software developed specifically for MR-application scenarios.

Fig. 1.4 The fiducial marker on the index finger of the user, left, and the participants view of the trocar and MR instructions during the assembly task, right.

As a result of previous user studies in this user group [25, 27] the interaction method chosen for the MR system was voice control. The voice input is received

through the headset microphone and is interpreted by a simple voice recognition application based on Microsoft's Speech API (SAPI).

1.4.2 The user task

The participants were given instructions on how to assemble a trocar (see Figure 1.4). 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 correctly assembled 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 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 1.4).

Fig. 1.5 A participant in the user study wearing the MR system and following instructions on how to assemble the trocar.

As described above, instructions on how to put together a trocar are normally given on the spot by more experienced operating room (OR) nurses. Creating the MR instructions was consequently somewhat difficult as there are no standardized instructions on how to put a trocar together. Instead we developed the instructions based on the instructions given by the OR nurse who regularly gives the instructions at the hospital. This ensures some realism in the task. The nurse was video recorded while giving instructions and assembling a trocar. The video was the basis for the sequence of instructions and animations given to the participants in the study. An example of the instructions and animation can be seen in Figure 1.4. Figure 1.5 shows a participant during the task.

Before receiving the assembly instructions the participants were given a short introduction to the voice commands they can use during the task; *OK* to continue to the next step, and *back* or *backwards* to repeat previous steps.

1.4.3 Participants and procedure

As the approach advocated in this chapter calls for real end users the selection of participants was limited to professional medical staff. Twelve professional (ages 35 – 60) operating room (OR) nurses and surgeons at a hospital took part in the study. As medical staff, the participants were all familiar with the trocar, although not all of them had actually assembled one prior to this study. None of them had previously assembled this specific trocar model. A majority of the participants stated that they have an interest in new technology, and that they interact with computers regularly, however few of them had experience of video games and 3D graphics. The participants were first introduced to the MR system. When the head mounted display and headset were appropriately adjusted they were told to follow the instructions given by the system to assemble the device they had in front of them. After the task was completed the participants filled out a questionnaire about their experience. The participants were recorded with a digital video camera when they assembled the trocar. During the task, the participants' view through the MR system was also logged on video. Data was collected both through direct observation and through questionnaires.

The observations and questionnaire were the basis for a qualitative analysis. The questionnaire consisted of 10 questions where the participants could answer freely on their experience of the MR system. The questions related to overall impression of the MR system, experienced difficulties, experienced positive aspects, what they would change in the system and whether it is possible to compare receiving MR instructions to receiving instructions from a teacher.

1.4.4 Results of the study

All users in this study were able to complete the task with the aid of MR instructions. Issues, problems or comments that were raised by more than one participant have been the focus of the analysis. The responses in the questionnaire were diverse in content but a few topics were raised by several respondents and several themes could be identified across the answers of the participants. None of the open ended questions were specifically about the placement of the marker, but the marker was mentioned by half of the participants (six of the twelve) as either troublesome or not functional in this application:

“It would have been nice to not have to think about the marker¹.” (participant 7)

Concerning the dual modality function in the MR instructions (instructions given both aurally and visually) one respondent commented on this as a positive factor in the system. But another participant instead considered the multimedia presentation as being confusing:

¹ All participant excerpts are translated from Swedish to English.

“I get a bit confused by the voice and the images. I think it’s harder than it maybe is.”
(participant 9)

Issues of parallax and depth perception are not a problem unique to this MR application. It is a commonly known problem for any video-see-through system where the cameras angle are somewhat distorted from the angle of the users eyes, causing a parallax vision, i.e. the user will see her/his hands at one position but this position does not correspond to actual position of the hand. Only one participant mentioned problems related to this issue:

“The depth was missing.” (participant 7)

A majority among the participants (eight out of twelve) gave positive remarks on the instructions and presentation of instructions. One issue raised by two participants was the possibility to ask questions. The issue of feedback and the possibility to ask questions are also connected to the issue of the system being more or less comparable to human tutoring. It was in relation to this question that most responses concerning the possibility to ask questions, and the lack of feedback were raised. The question of whether or not it is possible to compare receiving instructions from the MR system with receiving instructions from a human did get an overall positive response. Four out of the twelve gave a clear yes answer and five gave more unclear answers like:

“Rather a complement; for repetition. Better? Teacher/tutor/instructor is not always available – then when a device is used rarely – very good.” (participant 1)

Several of the respondents in the yes category actually stated that the AR system was better than instructions from a teacher, because the instructions were objective in the sense that everyone will get exactly the same information. When asked about their impressions of the MR system, a majority of the participants gave very positive responses:

“Very interesting concept. Easy to understand the instructions. Easy to use.” (participant 3)

Others had more concerns:

“So and so, a bit tricky.” (participant 6)

One question specifically targeted the attitude towards using MR in their future professional life and all participants responded positively to this question. This is perhaps the most important result, and also the most obvious from a developer point of view – if people from the target user group actually state that they want to use it in their work the system has a much higher promise than a perfectly good system which the target users have not requested or do not consider useful [5].

1.5 The Second Case Study – a Collaborative MR application

Collaborative work has been studied extensively in many different research domains, from sociological, psychological as well as organisational perspectives.

Technological tools which aid collaboration have also been developed within the broad range of research on computer supported collaborative work, such as decision support systems combined with teleconferencing systems.

Virtual environments have been used as tools for training and simulating collaborative work (for instance the CAVE system and the Virtual Workbench [10]), but few, if any, systems have actually been aimed for use in real crisis management situations. When personnel from different organisations work together under stress, as in many crisis situations, there is always a risk that misunderstandings emerge due to differences in terminology or symbol use. In linguistics, it is common knowledge that time has to be spent on establishing a 'common ground', or a basis for communication, founded on personal expectations and assumptions between the persons communicating with each other [4, 18]. Thus, providing means to facilitate establishing a common ground are important for efficient collaboration.

In addition to this, there are situation-specific problems that emerge in collaborative command and control tasks. Such tasks often circle around a shared representation of the current activities, as in the case of a situational map. Most organisations involved in such tasks, like the military or rescue services, have developed a library of symbols that can be utilized for representing units and events. Problems arise when representatives from different organisations are to work together, since they are used to working with their own, organisation-specific, symbols and conventions. This means that time has to be spent explaining and negotiating meaning when jointly creating and manipulating a shared representation, a tedious task to undertake when there is little time, as for example in the case of forest fire-fighting in, or close to, urban areas.

Furthermore, for each organisation there is information that is only interesting for the representatives from that organisation. Thus, commanders from different organizations need personalized views of the same situational map, for instance using MR.

1.5.1 Equipment used in the study

A multi user collaborative MR application was developed through the iterative design process described previously in this chapter.

The MR system was an early high fidelity prototype, which allowed the users to interact with virtual elements. It includes a Z800 3DVisor from eMagin² integrated with a firewire camera. The Mixed Reality system runs on a 2.10GHz laptop with 3 GB RAM and a 128 MB NVIDIA GeForce 8400M GS graphics card. In order to interact with the MR system the users had a joystick-like interaction device allowing them to choose objects and functions affecting their view of the digital map (see Figure 1.6).

² <http://www.3dvisor.com/>

Fig. 1.6 Interaction device, left, and the users display, right, with symbols and pointing used in the collaborative MR application.

1.5.2 The user task

The study was conducted with the purpose of evaluating the MR system and application design in order to improve it as a tool for collaboration between organisations. It is not possible to conduct experiments in real fire-fighting situations. Instead, to create a realistic study, we used a scenario where the groups had to collaborate, distribute resources and plan actions in response to a simulated forest fire and other related or non-related events. In the scenario the participants act as they would as on-scene commanders in their respective organisation. This means that they together have to discuss the current situation and decide how to proceed in order to fight the fire, evacuate residents, redirect traffic, coordinate personnel as well as dealing with any unexpected events that may occur during such incidents. The MR system was used as a tool for them to see and manipulate their resources and as a way to have an overall view of the situation (see Figure 1.6).

In order to create a dynamic scenario and realistic responses and reactions to the participants decisions in the final study, we used a gaming simulator, C3 Fire [13]. The simulator was run in the background by the research team (see Figure 1.7), where one member inserted information into the gaming simulator, for instance, that several police cars has been reallocated to attend to a traffic incident. The exercise manager acted as a feedback channel to the participants in order for them to carry out their work. For instance, when the reallocated police cars had reached their new destination the exercise leader returned with information to the participants. Other examples of information from the gaming simulator are weather reports, status of personnel and vehicles, the spread of the fire etc.

Fig. 1.7 The simulated natural setting (a helicopter base), left, and the gaming simulator that was controlling the experiment, right.

1.5.3 Participants and procedure

The MR application was evaluated in a study where nine groups with three participants in each group used the system in a simulated scenario of a forest fire. To promote realistic results, the participants were representatives from the three organizations in focus; the fire department, the police and the military. All participants were novices to the application and scenario, none of them were involved during the developing phase of the application. The setting was at a military helicopter base (see Figure 1.7).

The application was designed around an exercise in which the participants, one from each organisation (the fire and rescue services, the police department and the military), had to interact and work together to complete tasks in a dynamic scenario. The exercise was observed and the participants also answered questionnaires pertaining to the MR system design, and finally a focus group discussion was held where the participants could reflect on and discuss a series of topics relating to the exercise, the scenario and the MR system. The tools used for data gathering were mainly qualitative; observation, focus groups and questionnaires.

1.5.4 Results of the study

The results from the previously described design phase studies were very useful for improving the application scenario as well as the system. The head mounted display was a big improvement which allowed the participants to move around and interact more freely. The new interaction device was also appreciated and the participants found it very easy to use and quick to learn.

The added possibility to see hand gestures such as pointing on the digital map has simplified the interaction considerably and also resulted in a more natural interaction and better communication between the participants. In the redesigned application the participants had exactly the same view allowing them to alter their personal

image but still seeing the same map, and not as previously an organisation specific map. As noted by one of the participants during a group discussion:

“A common picture, everything is better than me telling someone what it looks like...you need to see the picture and not to hear my words. “ (pilot, pre-study workshop)

The main results of the study are not limited to the actual evaluation of the MR system and the application in itself. Even more importantly, the results are an indication that the naturalistic approach brought out comments that would be impossible had the task not been a relatively realistic task, and the participants not been professionals with experience from these types of tasks.

“What I would want, it’s a bit reactionary but I would like to have only a small red dot with an identifier on the unit that I have. I think that the vehicles were too big and that’s charming and nice visually but I would prefer working with tactical units and then add their designation, regardless of what vehicle it is, what it looks like I don’t care...That would increase my efficiency in this system even though it will be a more dull environment...” (fire department, day 2)

This quote illustrates the importance of real participants, iterative design and a holistic approach. As a representative from the fire department he is not impressed by the neat 3D graphics vehicles, instead he prefers tactical units. As he has been using the system in a close to natural setting he also provides additional information on how to make the interface more efficient (using red dots). The tactical unit can, as the participant suggests, be used for each individual organisation, however as another quote illustrates, the participants do not necessarily need to see more than a 3D object of the other organisations information:

“It’s good, maybe more ‘nice’ than ‘need’, but I think it’s good because it’s quick for those who are not familiar. If I would put up a lot of symbols otherwise maybe no one would know, because everyone has different symbols, but you see quickly what is a fire truck and what is a tanker, everybody understands what a helicopter is rather than me drawing if it is a medium heavy or heavy, you don’t know if that’s a transport helicopter or...so if you have an image it is easier” (pilot, day 1)

As a lesson learned the design could be enhanced by making the symbols on an individual level, meaning that the symbols representing other organisations could be pictorial objects, while the own organisations’ objects are represented in much more detail, as tactical units with identifiers rather than as visual 3D objects.

It is also important to note that to do a fair evaluation of the MR system the participants need some training. In any organisation a new type of technology or system is not expected to function perfectly on the very first day of use. This study was the first time any of these participants had any contact with MR technology and therefore the participants did find the second session with the MR system easier:

“I thought it went very well the last time. We are beginners and we had learned the talk, we had learned to point on the map. My opinion is that we passed” (fire department, day 2)

So basing a system evaluation on first time user experiences may not be very relevant since in real life people are actually trained, and after some time the use of the system will most likely be considerably more efficient and meaningful.

Another result of the study (including the pre-user study workshops) was the participants ability to see what possibilities MR technology hold for their professions. One organisation remarked that MR would be very useful as a tool to distribute situational pictures to command and control functions elsewhere in the organisational structure. This means that the command and control function on the field has the opportunity to convey what's going on to the command and control out of the field, that is, functions in the organisational hierarchy with responsibility for general planning of the operation. This, however, does not affect the purpose of the use of MR as a collaborative tool as this would merely be a form of information transmission or distribution rather than part of an ongoing collaborative task such as the one studied here. One participant saw the possibilities of distributed collaboration between the organisations, where one organisations' view can be shared with others while discussing the problem at hand and being able to point in each others field of view regardless of where the actors are situated geographically. This possibility was greatly appreciated:

“So you can be completely distributed but still see the hands of people? That's fantastic.”
(fire department, pre-study workshop)

The participants did not only appreciate the new design, it also gave them ideas on how to further develop the MR system and see the potential of its following applications. Even though the participants in many respects are very positive towards the new technology and the possibilities it may hold for their organisation, they are also realists and they are fully aware of the demands placed on new technology in their natural environment:

“I only have a hard time seeing it in a real large fire in reality. In a situation where you are supposed to practice like we have done here, where you practice making decisions, where you practice cooperating, where you are supposed to get the picture of what you need to talk about, and what decisions have to made, then it is perfect. But take this out in the forest, and then run out and give directions to someone else where I need a paper map anyway because he's not supposed to hang around, he's supposed to go there and he needs to know what to do, he needs to bring an operational picture when he leaves, then I don't see it yet.” (pilot, day 2)

The main goal of the study was to evaluate the potential of the technology and not to evaluate a finished design, and as such the study was successful. Even though the MR system is not yet final in design and implementation the professional opinions and comments received through this study are imperative for future design and implementation. The results also clearly show the importance of an iterative design process of MR applications.

1.6 Discussion

Both MR studies presented in this chapter share two qualities that make them stand out in comparison to most current MR-research; all users were professionals, and an

iterative design process with end user involvement was used. Furthermore, the underlying theoretical approach utilized was cognitive systems engineering which provides a view on design that differs from traditional cognitive theories, in some important respects. Cognitive systems engineering postulates that the user and his/her tools should be seen as a joint system, acting with a purpose. Most traditional human computer interaction study methods, which are the basis for most usability studies in the MR domain, are based on theories of cognition. They are therefore also based on a decomposed view of humans and the tools they use.

To apply a decomposed view of a human and a MR-system is in many respects a difficult task since the actual purpose of the MR system is to manipulate the perception of the user.

Another important aspect is that the designs used in the two studies have, as far as possible, tried to free themselves from design principles based on desktop systems. Already in the early stages of the first study, we discovered that such principles have little value for designing MR applications. The reason for this lays in the fact that presentation and interaction with MR objects differ greatly from systems based on the desktop metaphor.

The method used in these case studies included a long development phase with several design iterations and end user representatives were involved throughout the entire process. This method is invaluable in terms of the final outcome – these experts found issues in the system that would be very difficult for non-experts in the field to spot. For instance many user development methods involve students during the design phases. However, for many domains, such as health care and civil service, involving expert users is imperative. The participants in the design iterations gave invaluable insight into what information is actually useful in the field, and what information is not as important. This gives meaningful and relevant feedback on what changes should be prioritized. Real expert users also have knowledge and experience allowing them to interpret the information in context. For instance, the information displayed in relation to objects in the collaborative task is useful to professionals in the field but may appear meaningless to someone without domain expertise.

Real end users in the evaluation also provided new insights on the use of MR systems and future research and development. Where many other studies result in findings where tracking optimisation and better resolution displays are the most important issues, the results from our end user case studies actually show that the resolution and tracking are of minor importance to the end user, c.f. Gabbard and Swan, 2008 [11]. What is of importance however is the interaction possibilities and the content of the presented information. One negative aspect of end user involvement is of course the risk of bias – if the user is part of the development team there is the risk that no matter what the design result is they will be positive. However, this can to some extent be controlled in that the users actually taking part in the study and evaluation of the application are not involved in the development phase. In our studies the participants in the end studies had no previous involvement in the project and were informed that the results of the study would be used solely for research and development purposes, and that their individual performance and opinions were

confidential. Despite this there can of course be reasons out of our control for them to be positively or negatively biased as representatives from the end user group. No matter what study, when involving users the potential bias in one way or another will be present, but the advantages and possible gain from their involvement definitely outweighs the potential risks.

1.7 Conclusions and future direction

New designs, based on novel technologies, like the ones presented in this chapter naturally lacks in maturity. There are child diseases in terms of technical limitations, various bugs due to experimental software, and the mere fact that the users are not acquainted with the way of interacting. This leads to problems in applying traditional evaluation methods that for example focus on efficiency. Instead, when introducing new systems, like MR, in an activity, user acceptance and perspectives are crucial evaluation criteria. The overall results from both case studies presented in this chapter show a system that the participants like rather than dislike, although both need specific improvements before they can be implemented into the context of use. Both user groups clearly see a future for the MR technology in their work domains. By allowing professional users to test the applications new ideas are generated, both on behalf of the users and the designers. The 'envisioned world problem' described above certainly manifests itself. However, the word 'problem' may be an inappropriate notion. When working with novel systems like these, we should perhaps rather talk about possibilities. Even if the actual use differs from the envisioned, every iteration has so far provided very important input that definitely has led to large improvements and innovation.

Interactivity is an important part of direct manipulation user interfaces and also seems to be of importance in an MR system of the kind investigated in these studies. In the first case study, a couple of the participants who responded negatively on the question regarding comparability between the MR system instructions and human instructions, motivated their response in that you can ask and get a response from a human, but this MR system did not have the ability to answer random questions from the users. Adding this type of dialogue management and/or intelligence in the system would very likely increase the usability and usefulness of the system, and also make it more human-like than tool-like [29]. In the second study we found that showing real time status of vehicles and resources is one important output improvement. Connecting the MR interface with other technological systems already in use in the organizations, GPS for instance, would enhance the MR system. We also found that the interaction device could be used more efficiently if buttons and menu items were organised differently.

We believe that our iterative design method gave us MR systems that were appreciated by our subjects. However, the experiences from our case studies also show the need for further research on user perspectives on interaction and output in MR systems.

Acknowledgements This research is funded by the Swedish Defence Materiel Administration (FMV). The MR system was developed in close cooperation with XM Reality AB. We are deeply indebted to all the participants in our studies who volunteered their time and expertise to these projects.

References

1. Azuma, R.: A survey of augmented reality. *Presence* **6**(4), 355–385 (1997)
2. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. *IEEE Computer Graphics and Applications* **21**(6), 34–47 (2001). URL <http://computer.org/cga/cg2001/g6034abs.htm>
3. Billinghurst, M., Kato, H.: Collaborative augmented reality. *Communications of the ACM* **45**(7), 64–70 (2002)
4. Clark, H.H.: *Using Language*. Cambridge University Press, Cambridge (1996)
5. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* **13**(3) (1989)
6. Dekker, S., Hollnagel, E.: Human factors and folk models. *Cogn. Technol. Work* **6**(2), 79–86 (2004). DOI <http://dx.doi.org/10.1007/s10111-003-0136-9>
7. Dünser, A., Grasset, R., Billinghurst, M.: A survey of evaluation techniques used in augmented reality studies. Tech. Rep. Technical Report TR2008-02, Human Interface Technology Laboratory New Zealand (2008)
8. Engeström, Yrjö, Miettinen, Reijo, Punamaki, Raij (eds.): *Perspectives on Activity Theory (Learning in Doing Social, Cognitive and Computational Perspectives)*. Cambridge University Press (1999)
9. Fiala, M.: Artag rev2 fiducial marker system: Vision based tracking for ar. In: *Workshop of Industrial Augmented Reality*, Vienna Austria (2005)
10. Fuhrmann, A., Löffelmann, H., Schmalstieg, D.: Collaborative augmented reality: Exploring dynamical systems. In: R. Yagel, H. Hagen (eds.) *IEEE Visualization '97*, pp. 459–462. IEEE (1997)
11. Gabbard, J.L., Swan II, J.E.: Usability engineering for augmented reality: Employing user-based studies to inform design. *IEEE Transactions on Visualization and Computer Graphics* **14**(3), 513–525 (2008)
12. Glesne, C., Peshkin, A.: *Becoming qualitative researchers: An introduction*. White Plains, NY: Longman (1992)
13. Granlund, R.: *Web-based micro-world simulation for emergency management training*. Future Generation Computer Systems (2001)
14. Grasset, R., Lamb, P., Billinghurst, M.: Evaluation of mixed-space collaboration. In: *ISMAR '05: Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 90–99. IEEE Computer Society, Washington, DC, USA (2005). DOI <http://dx.doi.org/10.1109/ISMAR.2005.30>
15. HITLAB: <http://www.hitl.washington.edu/artoolkit/> (2007)
16. Hollnagel, E., Woods, D.D.: Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies* **18**(6), 583–600 (1983)
17. Hollnagel, E., Woods, D.D.: *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Press, Boca Raton, FL (2005)
18. Klein, G., Feltovich, P.J., Bradshaw, J.M., Woods, D.D.: Common Ground and Coordination in Joint Activity, vol. *Organizational Simulation*, chap. 6, pp. 139–184. John Wiley & Sons (2005). [Http://dx.doi.org/10.1002/0471739448.ch6](http://dx.doi.org/10.1002/0471739448.ch6)
19. Liao, Z., Landry, R.: An empirical study on organizational acceptance of new information systems in a commercial bank. In: *HICSS (2000)*. URL <http://computer.org/proceedings/hicss/0493/04932/04932021abs.htm>

20. Lincoln, Y.S., Guba, E.G.: *Naturalistic Inquiry*. Sage (1985)
21. Livingston, M.A.: Evaluating human factors in augmented reality systems. *IEEE Computer Graphics and Applications* **25**(6), 6–9 (2005). URL <http://doi.ieeecomputersociety.org/10.1109/MCG.2005.130>
22. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems* **E77-D**(12) (1994)
23. Neisser, U.: *Cognition and reality*. W.H. Freeman, San Francisco (1976)
24. Nielsen, J.: *Usability Engineering*. Academic Press (1993)
25. Nilsson, S., Johansson, B.: User experience and acceptance of a mixed reality system in a naturalistic setting - a case study. In: *ISMAR*, pp. 247–248. IEEE (2006). URL <http://dx.doi.org/10.1109/ISMAR.2006.297827>
26. Nilsson, S., Johansson, B.: Fun and usable: augmented reality instructions in a hospital setting. In: B. Thomas (ed.) *Proceedings of the 2007 Australasian Computer-Human Interaction Conference, OZCHI 2007, Adelaide, Australia, November 28-30, 2007, ACM International Conference Proceeding Series*, vol. 251, pp. 123–130. ACM (2007). URL <http://doi.acm.org/10.1145/1324892.1324915>
27. Nilsson, S., Johansson, B.: Acceptance of augmented reality instructions in a real work setting. In: M. Czerwinski, A.M. Lund, D.S. Tan (eds.) *Extended Abstracts Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008, Florence, Italy, April 5-10, 2008*, pp. 2025–2032. ACM (2008). URL <http://doi.acm.org/10.1145/1358628.1358633>
28. Patton, M.Q.: *Qualitative Evaluation and Research Methods*, 2 edn. Sage Publications, Thousand Oaks, California (1990)
29. Qvarfordt, P., Jönsson, A., Dahlbäck, N.: The role of spoken feedback in experiencing multimodal interfaces as human-like. In: *Proceedings of ICMI'03, Vancouver, Canada (2003)*
30. Reichardt, C., Cook, T.: Beyond qualitative versus quantitative methods. *Qualitative and quantitative methods in evaluation research* pp. 7–32 (1979)
31. Schmalstieg, D.: Rapid prototyping of augmented reality applications with the studierstube framework. In: *Workshop of Industrial Augmented Reality, Vienna Austria (2005)*
32. Shneiderman, B.: *Designing the User Interface*, 3 edn. Addison Wesley (1998)
33. Suchman, L.A.: *Plans and Situated Actions*. Cambridge University Press, Cambridge (1987)
34. Sutherland, I.E.: A head-mounted three-dimensional display. In: *AFIPS Conference Proceedings*, vol. 33, pp. 757–764 (1968)
35. Swan II, J.E., Gabbard, J.L.: Survey of user-based experimentation in augmented reality. In: *1st International Conference on Virtual Reality, Las Vegas, NV (2005)*
36. Träskbäck, M.: Toward a usable mixed reality authoring tool. In: *VL/HCC*, pp. 160–162. IEEE Computer Society (2004). URL <http://doi.ieeecomputersociety.org/10.1109/VLHCC.2004.60>
37. Tushman, M.L., O'Reilly III, C.A.: Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review* **38**(4), 8–30 (1996)
38. Vygotsky, L.S.: *Mind In Society*. Harvard University Press, Cambridge (1978)
39. Woods, D.D., Roth, E.M.: Cognitive engineering: human problem solving with tools. *Hum. Factors* **30**(4), 415–430 (1988)