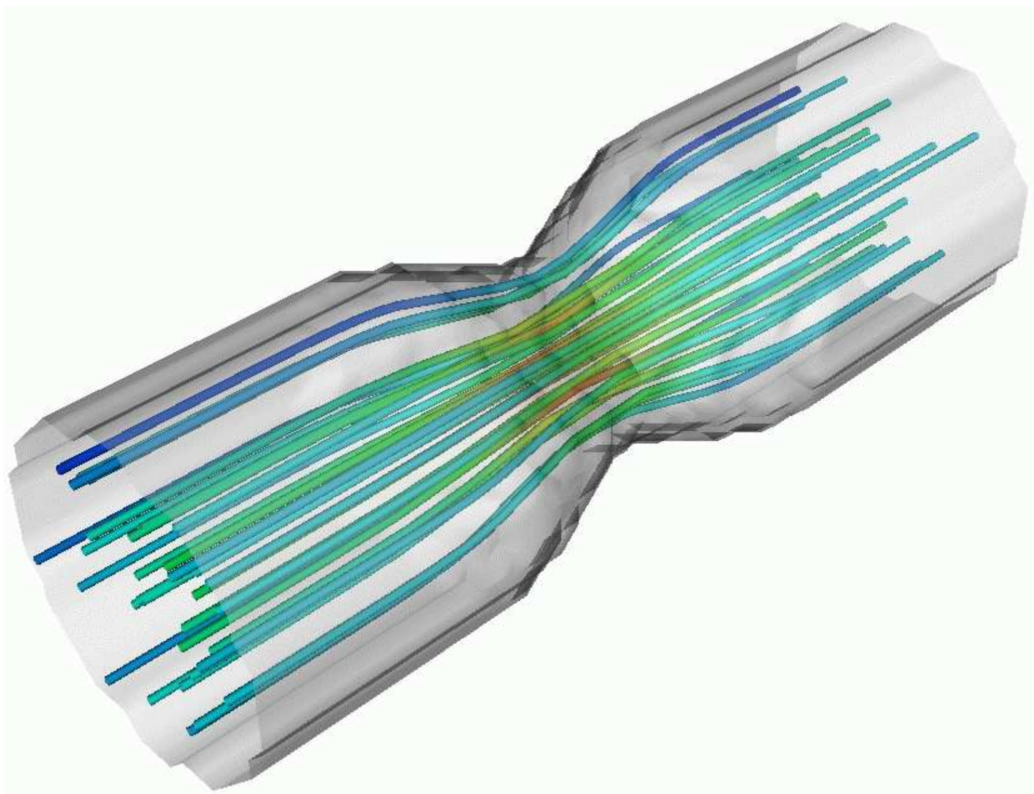
Usability Evaluation And Context Analysis To Aid The Development Of Virtual Reality Applications

applied to the
Virtual Radiology Explorer



H.S.M. Cramer

In partial fulfilment of
the requirements for the
Master's degree in Social Science Informatics

Faculty of Science
University of Amsterdam
The Netherlands

Preface

This thesis describes the result of a study performed in partial fulfilment of the requirements for a Master's degree in Social Science Informatics at the University of Amsterdam. This study examines usability issues development teams of Virtual Reality (VR) systems currently encounter and studies the application of context analysis and usability evaluation in VR system development. A case study towards this goal has been conducted into usability of a medical VR system; the Virtual Radiology Explorer, in a clinical context of use.

Parts of this thesis have been accepted for publication / presentation in:

H.S.M. Cramer, V. Evers, E.V. Zudilova V., P.M.A. Sloot (2004) Context analysis to support development of virtual reality applications, *Journal of Virtual Reality*, 7(3), 177-186.

H.S.M. Cramer, V. Evers, E.V. Zudilova V., P.M.A. Sloot (2004) Heuristic evaluation and context analysis to aid the development of a simulated vascular reconstruction system *Proceedings of Virtual Reality Design and Evaluation Workshop*, 22-23rd January, 2004, Nottingham, United Kingdom, Proceedings available on CD; ISBN 0 85358 123 1.

Sponsoring

The VRE research is partially sponsored by the EU CrossGrid Project IST-2001-32243 and the Token 2000 project Distributed Interactive Medical Exploratory for 3D Medical Images.

Front page illustration

The illustration on the front of this thesis is an mage of simulation of blood flow in a stenotic part of an artery (Belleman, 2003, image created by Hans Ragas).

Abstract

Virtual Reality (VR) systems have been available for years, but are not widely used in contexts outside research laboratories. Apart from budgetary concerns, this might be partially caused by utility and usability problems of VR applications. It is argued here that for a Virtual Reality application to be actually used, usability and the fit with the context in which a system could be used have to be addressed in system design and development along with utility questions.

This thesis focuses on the development of guidelines for VR usability and the applicability of contextual analysis to inform the development of VR systems. For this purpose a medical Virtual Reality system, the Virtual Radiology Explorer (VRE), is evaluated. The VRE aims to assist radiologists and vascular surgeons in diagnosis and treatment planning for vascular disorders. It offers visualization of arteries in 3D, blood flow simulation and capabilities to try out virtual bypasses in an semi-immersive virtual environment. The development of the VRE prototype centred around technological feats, rather than on usability and implementation in an actual context of use. At a later stage, interest in actual implementation of the system has resulted in a need for information about the clinical context in which the VRE might be used and ways to improve usability of the VRE.

After evaluation of existing literature, an exploratory study was carried out to gain more insight in potential usability problems of the VRE and its fit within current work practices regarding vascular disorders in radiology and vascular surgery departments in the Netherlands. First, a heuristic usability evaluation has been conducted to gain insight in the problems users may encounter while working with the current VRE prototype. This evaluation has resulted in identification of usability problems of the prototype and VR technology in general. An additional result however was the realization that a closer look at context of use and desired utility of the system would be necessary. An exploratory context study was subsequently conducted. This study evaluated the processes of diagnosis and

treatment planning of vascular disorders within the daily work environment of radiologists and vascular surgeons. Semi-structured interviews were coupled with an ethnographic approach to requirements engineering to analyse the context of use. This resulted in more insight in functional and non-functional requirements for VR systems that aim to support diagnosis and treatment planning processes. The findings have indicated that contextual analysis can be a useful way to inform design of a VR application.

Results show that adjustments to both utility and usability are necessary to make the VRE system suitable for a clinical environment. It was discovered that some functionality (e.g. blood simulation and 3D visualization options) did not fit user requirements and has to be re-examined in order to provide a perceived value over existing systems. Usability issues for the VRE include problems concerning interaction technologies, such as voice recognition and tracking of user position, which were bound to suffer from inaccuracies and are also not particularly suited to current clinical contexts. Inadequacies can be identified for available manipulation devices as well, including in the area of ergonomics. Most notably however, some major usability problems seem to result from the choice for using the VR paradigm for the VRE application. This while the virtual world on offer does not seem to offer much additional value to this particular application. Whether fully immersive VR and stereoscopic images are currently wanted and needed in a Dutch clinical context remains questionable.

Context analysis can help decide whether a VR application is suitable and usable in a particular context of use. However, usability problems surrounding VR systems and input technology need further attention in general VR research. Concise guidelines on design choices for VR systems and effective specialized evaluation methods to investigate usability of VR systems are still scarce. It is yet unclear when applying the immersive virtual environment paradigm is useful or whether offering stereoscopy is useful in a particular situation. Walkthrough type applications (such as architectural applications) which aim to provide the user with a sense of presence might benefit more from an immersive environment than applications that aim to provide insight in the structure of an virtual object (such as the VRE). Further research is needed to assess for which user groups, tasks and contexts a Virtual Reality system would be a suitable option and what type of environment and sense of immersiveness would be most supportive.

Table of contents

1. Introduction

Virtual Reality as a computing paradigm has been in existence for a number of years. However, achieving usable and affordable VR systems is still a challenge and widespread proliferation of VR systems has by no means occurred. Virtual reality can at times even be viewed as ‘a solution looking for a problem’ (Bowman et al., 2002); it is unclear for what type of situations Virtual Reality is actually useful and what problems can be solved with the VR paradigm. Virtual Reality systems have been presented as the promise of an ultimate interface, offering natural interaction in a convincing alternate reality (see Gobbetti and Scateni, 1999, Chesher, 1994). However, VR systems are still largely perceived as technological experiments and serious usability problems can be identified. Actually developing VR systems that can and are used outside of the laboratory is challenging, both technically and in respect to human factors and usability issues.

It has been recognized that user needs should play a central role in the development of virtual environments if they are to be used in a real-life context (e.g. Crabtree et al., 2002), an insight that has existed for general system development for years (e.g. Hughes et al., 1995). However, in VR research the attention has often been put on technological experimentation, instead of use of developed VR systems by prospect users. Usability guidelines specific to VR systems are scarce and information on how to optimise a system for a prospect context of use is not readily available.

This thesis will not focus on the complex technical aspects of VR development; many researchers already address these. Studies actually involving context analysis to investigate whether a VR system could actually effectively fulfil prospect users’ needs is scarcer. Focus in this thesis will be on the human-computer interaction and usability issues surrounding VR systems, user requirements for these systems and fit in a non-experimental context of use. It is argued in this thesis that developing a system that will be used in a context outside an experimental setting requires insight on usability issues and extensive information on the prospect context of use of the system at hand.

This thesis describes a case study that has been conducted to explore usability and contextual issues that might affect future use of a VR system. The study described revolves around the usability evaluation and context analysis of an experimental, multi-modal Virtual Reality application; the Virtual Radiology Explorer (VRE). The VRE is a medical VR system that aims to support diagnosis and treatment planning for aortic abdominal vascular disorders. The prototype is an experimental set-up offering 3D visualization of medical scans and simulation of blood flow in a virtual environment. The user can review 3D reconstructions of scan data of aortic vessels, try out a virtual aortic bypass and assess its effects on blood flow. The VRE has been developed largely as a technological test-bed, as many VR applications currently are. Focus has been mainly on technological feats and not on actual application of the system in a non-experimental environment. Usability of the prototype and its functionality might subsequently not be ideal for prospect users. There is now however an interest in finding out whether actually implementing the VRE in a clinical context would be possible and desirable. A usability evaluation and context analysis that have been conducted to gather information about the requirements for such a system are discussed in this thesis. Usability and utility issues that might come into play when implementing the system for actual end-users are extensively discussed.

With the VRE case study insight is to be gained into usability issues for VR development in general, the contextual criteria that might play a role in usability in an actual context of use and criteria for the choice for type of (VR) environment. It will try and identify usability challenges that are encountered in developing a VR system for which existing literature cannot yet provide an answer. The intention here is not necessarily to promote use of VR systems, but to identify problems that counter use of VR systems and try and aid improvements in usability of VR systems in situations where they are applicable. The objective is to give guidelines for development of VR systems in general as well as for VR systems for medical diagnosis and planning in specific.

1.1. Problem statement

While literature on usability problems of VR systems is available, little attention is given to the questions whether a VR system would actually fit a prospect context of use and which (unresolved) issues are encountered during development of a system actually meant for use outside the laboratory.

This thesis aims to provide insight on the following questions:

- What utility and usability issues exist for utilization of VR systems in a real world context?
    - Which usability issues make developing usable VR systems challenging?
    - Do VR systems provide sufficient added value to warrant their use in spite of potential drawbacks?

The VRE system has been chosen as a case study to gain more insight in these issues. Specifically in the case study, the ways are studied, in which context of use affects use of the VRE and what requirements exist for utility and usability of the system. The above questions have been adapted for the case study. The specific research questions tailored to systems such as the VRE, that aim to support vascular diagnosis and treatment planning are introduced in section 3.

For general VR research the study aims to gain insight in how context of use affects utility and applicability of VR systems and in this way develop guidelines that could support VR development. With this study insight is also expected to be gained on the applicability of available usability evaluation methods and methods for analysis of a context of use to VR development.

1.2. Thesis outline

This thesis is structured as follows:

Chapter 2 discusses the main concepts from literature that play a role in this study; focus is on Virtual Reality, usability and context of use. The chapter will first give a short history of Virtual Reality system development and will discuss various interpretations of the concepts of Virtual Reality, usability and context of use. The concepts of immersion and presence are presented, as they are central to many studies surrounding VR. Various types of VR systems will be discussed, illustrating the wide choice of systems available and the distinction to other types of (computer) systems.

Chapter 3 introduces the case study central to this thesis and the Virtual Radiology Explorer system will be further highlighted. The questions that have arisen of the VRE system in an actual clinical context concerning the utility of its functionality, usability and contextual constraints will be discussed in more detail. This chapter will also introduce a number of issues that come into play in developing medical VR systems in general.

Chapter 4 describes the methodology, and the rationale, of the usability evaluation and context analysis that has been performed to inform further development of the VRE. First the methodology followed of heuristic walkthrough evaluation of the VRE is elaborated upon. After which, the ethnographic context analysis that has been taken to explore the prospect context of use of the VRE is discussed.

Results of the usability evaluation and context analysis are given in Chapter 5. Problems resulting from the prototype stage of the VRE, interaction methods and problems resulting from the choice in this case for the VR paradigm are discussed. In the second part of the chapter a discussion follows about the observed trajectory followed in diagnosis and treatment planning for vascular disorders, the types of activities observed, scan use during the process characteristics of prospect users of the VRE and the social, physical and technical environment of systems deployed in this clinical context.

Chapter 6 provides a discussion of the results and methods of this study. First the chapter will explore the ways in which the findings of the study impact development of the VRE, after which the implications for VR system development in general are discussed. Considerations for similar research and limitations of the methodological approach of this study are given.

In chapter 7 provides the conclusion of this thesis. Conclusions that can be drawn from the described study are discussed. Open questions that have risen during this study and possibilities for future research will conclude this thesis.

2. Literature background

This section will shed light on a number of concepts from the existing literature that are explored in this thesis. It will elaborate on the paradigm of Virtual Reality, and the systems that might be called virtual environments. The concepts of immersion and presence are discussed, as they are central issues in many discussions in Virtual Reality research. Furthermore, usefulness and usability of a system are elaborated upon. Various ways are explored that can lead to usable systems. Special attention is given to usability guidance in literature for VR systems. The importance of considering a system's context of use in achieving usability is highlighted as well. A number of methods for requirements engineering that can be used to gain insight in the contextual constraints of a system are described. The section starts out with a discussion of the concept of Virtual Reality.

2.1. Virtual Reality

Virtual Reality systems have been available for a number of years; the following section describes the concept of VR and its history. It provides a background to the human factors issues at play in interaction with VR that can make developing a usable system difficult. Various taxonomies distinguishing VR systems from other graphical systems are explored, as well as several existing categories of VR systems. First a short history of the concept of Virtual reality is given.

2.1.1. Short history of Virtual Reality

The concept of Virtual Reality (VR), even if introduced by using other terms, dates from the 1960's. In 1962 Morton Heilig introduced his Sensorama 'reality simulator' (fig. 1). Sensorama presented its users with a simulated motorcycle ride. The system provided 3D visuals, stereo sound, wind, touch and smell. The images were not computer generated, but taped with cinematographic technology. The system essentially offered no interactivity, only playback of a multi-sensory experience. The Sensorama never became a commercial success, but it did generate interest in the vision of a multi-sensory experience (as discussed by Balaguer and Mangili, 1991and Belleman, 2003).

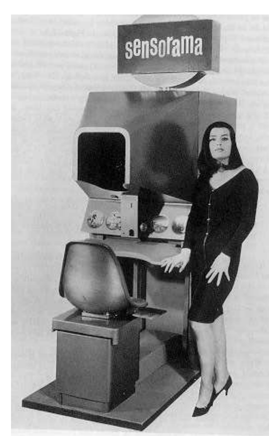


Fig. 1 Heilig's Sensorama, 1962.

The first system now considered an actual Virtual Reality system was envisioned in 1965 by Ivan Sutherland, (see Gobbetti and Scateni, 1999; Balaguer and Mangili, 1991; Slater and Usoh, 1994; Belleman, 2003). Sutherland presented the idea of immersing people in worlds generated by computers, instead of offering a view of reality itself. He wanted to present a view mimicking reality through simulation or moving beyond reality by bypassing laws of physics that exist in real life. Sutherland realized his vision in 1968 with his 'Ultimate Display' (fig. 2), where data visualization was combined with interaction. The Ultimate Display included a head mounted display with two cathode-ray tube displays mounted on a helmet. The user was presented with a stereoscopic view of 3D wireframe objects in a virtual environment. The 'sword of Damocles', a device attached to both ceiling and helmet, tracked the head position of the user. The view was adjusted according to the position of the user's head, thus achieving interaction.

Following Sutherland's pioneering work, research into virtual environments was conducted at amongst others NASA and MIT (see Balaguer and Mangili, 1991). However, Virtual Reality remained unknown to the general public until 1989, when the companies VPL and AutoDesk presented their VR systems at tradeshows. Both companies there showcased head-mounted displays and devices for interacting with virtual worlds (Chesher, 1994; Balaguer and Mangili, 1991; Bricken, 1990).

Fig. 2 Sutherland's Ultimate Display, 1968.

Since then, new VR applications, also called Virtual Environments (VE's), have been implemented in a variety of fields such as medicine, training, architecture and scientific visualization (see for overviews Balaguer and Mangili, 1991; Waterworth, 1999). What constitutes a VR system is however not always clear, the following section will try and define the concept.

2.1.2. Defining Virtual Reality

The introduction of VR has proposed a paradigm shift. The term has changed popular impressions of what technology can and should do (Bricken, 1990; Hand, 1994; Chesher, 1991). Computers offering VR are no longer only symbol processors, but can now create an alternative reality. Virtual worlds, including inhabitants, can be generated by computers and presented to the user, who in turn can interact with these worlds (Bricken, 1990). The term Virtual Reality itself was coined in the late nineteen-eighties by Jaron Lanier, the founder of VPL Research. Alternative terms had been proposed before then and have been introduced since. Marsh et al., (1998) sum up a few examples: artificial reality, spatial immersion and virtual presence (for those who consider the terms 'virtual' and 'artificial' principally contradictory to 'reality'). Virtual Environments (VE's) is the term commonly used for VR systems.

What Virtual Reality is exactly (and what it is not) is however not yet completely agreed upon. The label of Virtual Reality has been attached to a diversity of new ideas and technologies, taking on a variety of meanings. In both academic and popular publications the term VR has been used to refer to anything from email to a fully surrounding, multi-sensory environment (Bricken, 1990; Hand, 1994; Chesher, 1994).

Taking a narrow definition of VR, a VR system is one similar to Ivan Sutherland's Ultimate Display: a system that can present information to all the user's senses at a resolution equal to or greater than that he or she can discern so there is no way to tell that the artificial world is not real (Steed, 1993). Such a system does not yet exist, but its enabling technologies are under development. It needs to be noted that authors such as Hand (1994) warn of the effects on putting too much emphasis on technology in achieving such an ultimate goal of creating an alternative reality. He draws a parallel of VR to artificial intelligence, where solely trying to achieve strong AI and 'thinking machines' would lead to a dead end. Instead, VR technology can be used for practical means in developing systems that might not offer a truly convincing alternate reality, but are useful none the less. Indeed, an alternate reality might not be always needed or useful.

The goal of providing a convincing alternate reality does however occur in many of the definitions of VR given in literature. Ellis' concept of virtualization is defined as 'the process by which a human viewer interprets a patterned sensory impression to be an extended object in an environment other than that in which it physically exists' (Hullfish, 1996). Chesher (1994) defines the goal of VR technology as 'to fool people's senses into believing they are in the artificial environment'. Similarly, Bricken (1990) states that the term VR generally refers to the body of techniques that apply computation to the generation of experientially valid realities. He believes that the primary defining characteristic of VR is inclusion, being surrounded by an environment. Gobbetti and Scateni (1999) define the goal of VR as putting the user in the loop of a real-time simulation, immersed in a world that can be both autonomous and responsive to its actions. Hand (1994) defines Virtual Reality as the paradigm whereby we use a computer to interact with something "which is not real, but may be considered to be real while using it". Milgram (1994) states that the 'conventionally held view of a VR environment is one in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world'. Considering above definitions and by somewhat stretching their boundaries, text-based systems such as multi-user dungeons, still can be (and are) proclaimed Virtual Reality. However, focus in this thesis is not on such socially or mentally constructed 'virtual environments' or communities, but on virtual environments that offer actual 3D imagery of a virtual environment. The following section elaborates on frameworks that try and position such graphical VR systems in relation to other graphical systems.

2.1.3. Distinguishing Virtual Reality systems from other systems

A number of taxonomies have been developed to place VR in the wider context of graphical systems. One of these is Zeltzer's Autonomy-Interaction-Presence (AIP) cube (1992, fig. 3). Zeltzer's AIP cube is a three-dimensional taxonomy of graphical simulation systems. The axes of the cube represent Autonomy, Interaction or Presence, on a scale from 0 to 1. Autonomy is defined as the ability of the environment to act and react to simulated events. Interaction is the fidelity with which the environment deals with interactions between its participants both human and synthetic. Presence provides a rough (dimensionless) measure of the number and fidelity of available input and output channels. The point (1,1,1) is often agreed on as the ultimate VR (Hand, 1994; Kalawsky et al., 1999).

Zeltzer claims that any VR solution can be characterized with a position within the cube. Kalawsky et al., (1999) state that it is very difficult to characterize a VR system in this way, because of the lack of a clear definition for each axis. Measuring the three 'qualities', either objectively or subjectively using defined metrics is not yet addressed. Especially presence is difficult to specify and is a debated term. Kalawsky et al., further point out that Zeltzer associates the position of an application within the cube with task performance. There is a danger that the real performance controlling factors of a VR system will not be addressed by devising a measurement tool for only Zeltzer's three attributes of a VR system, let alone that the contextual factors influencing task performance are addressed in this way.

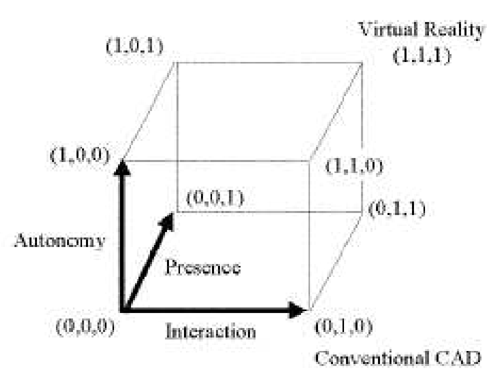


Fig. 3 Zeltzer's AIP-cube (Zeltzer, 1992).

Marsh et al., (1998) propose the use of a shared framework of VR to identify common features, attributes and concepts of VR; and to distinguish between VR and other computerized graphical systems and media that induce VR-related qualities such as presence and/or immersion (e.g. film). This framework is an effort to provide a collective understanding and terminology for VR.

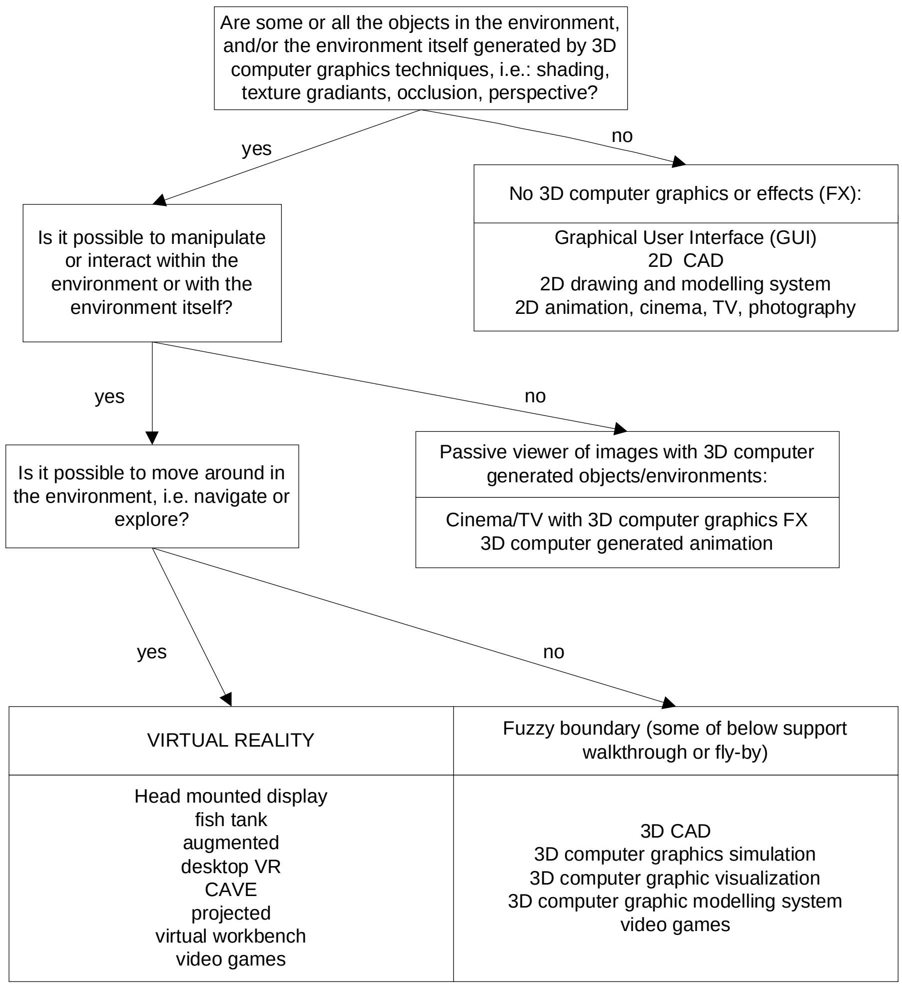


Fig. 4: Framework for Virtual Reality (Marsh et al., 1998, notation and terms changed for reasons of clarity and consistency when applicable)

Milgram's Reality-Virtuality continuum (1994, fig. 5) in turn tries to position systems on the continuum from a real environment, consisting of only real objects (watching a real world scene either via or not via an electronic display system) to a completely virtual environment. The continuum is meant to provide a deeper understanding of 'Mixed Reality' systems that can be placed somewhere between actual reality and Virtual Reality, but is often referred to when trying to define Virtual Reality itself. Examples of mixed reality systems are augmented reality systems, see-through head mounted displays, or systems in which for example the hand of the user is visible and used as a direct manipulation device.

Fig. 5: Simplified Milgram Virtuality Continuum (Milgram, 1994)

Milgram distinguishes Mixed Reality display systems from each other on three dimensions: Extent of World Knowledge ("How much do we know about the world being displayed? To what extent is this world modelled?"), Reproduction Fidelity ("How 'realistic' is the display?"), and Extent of Presence Metaphor ("What is the extent of the illusion that the observer is present within that world?"). Reality and Virtual Reality are on the ends of the scale on the first dimension of world knowledge; Virtual Reality offers a completely modelled world and a VR system thus has to contain all information on that world. The other two dimensions deal with the realism of the environment (or the illusion thereof); both a completely convincing Virtual Reality and reality itself can be found on the same end of these scales. These dimensions of reproduction fidelity and extent of presence are less suited to distinguishing between systems, nor are they meant to.

The frameworks outlined above can help to characterize VR systems, but are currently not equally suited to distinguishing VR from other graphical systems. Milgram's dimensions are not meant to distinguish between other graphical systems and VR. Rather the Milgram continuum is focused at characterizing the properties of mixed reality systems, which can be viewed as a specific subset of VR systems. Lack of definition and operationalization of the concepts in Zeltzer's cube make practical application in categorizing systems as VR systems difficult as well. It seems more practical to use the framework by Marsh et al. to determine if a system at hand can be considered a Virtual Reality system. Due to the explicit yes/no questions asked it is relatively easy to categorize a system at hand. An added advantage is that the framework also provides labels for other types of graphical systems and the framework is thus not limited to a yes/no answer to the question if a particular system is a VR system. The framework however is a starting point from the viewpoint of Marsh et al. and does not yet offer a concise definition. Taking into account the various definitions and frameworks discussed above, a working definition for the purposes of the study described in this thesis will be derived in the following section.

2.1.4. Definition of Virtual Reality for the purpose of this thesis

Various authors describe what they consider key concepts of VR that, when taking a closer look can be seen as inconsistent with the broad application of the term. Sutherland with his vision of the Ultimate Display introduced two of these concepts: immersion and complete sensory input and output interaction (Balaguer and Mangili, 1991; Gobbetti and Scateni, 1999). Complete sensory input and output involving all the user's senses have however not yet been achieved. While such things as haptic feedback are applied in VR systems, most research on VR still focuses on 3D graphical representation. Marsh (1998) defines the three central components to define all 3D VR systems: imagery (completely or in part 3D computer graphics), behaviour (the 3D computer generated items and environment behave like their real-world counterparts) and interaction (the user can manipulate and interact with the environment and objects, this includes navigation and exploration of the environment). It should be noted that the behaviour component does not exclude moving beyond reality and offering possibilities to the user not available in real life, such as flying or investigating a gigantic molecule from within. E.g. Bowman and Hodges (1999) take the stance that instead of replicating the real world, virtual environments should enhance the physical, cognitive and perceptual capabilities of the user. Completely mimicking real world interaction would be very difficult anyway and replication of the physical world can severely limit productivity.

In line with the above concerns, in the context of this thesis not all systems that have been proclaimed virtual environments are considered; so-called text-based virtual environments for example are excluded. Lanier's original definition of VR as "a computer generated, interactive, three-dimensional environment in which a person is immersed" (as reported in e.g. Gobbetti and Scateni, 1999) is somewhat more specific and in keeping with the purposes of this thesis. A VR system here will be defined as a system that offers computational simulation, 3D visualization and an artificial environment that the user can interact with in real-time. This definition takes in the three central components of VR systems as defined by Marsh et al., (1998) of imagery in 3D, behaviour of virtual objects simulates that of real objects and interaction.

Still, this definition of VR encompasses a wide array of systems. Further insight in the categories of VR is needed to provide system developers with the possibility to make an informed choice for a particular type of system; one that fits user's task goals, usability concerns and contextual constraints. One of the questions for a choice of VR system is the offered grade of immersiveness. Immersion and presence are central concepts in many VR related studies and the next section will accordingly discuss these concepts.

2.1.5. Immersion and presence

The central issue that seem to put VR apart from other technologies is the possibility to offer the experience of immersion and presence (Marsh et al., 1998). Both terms relate to the 'sense of being in the virtual environment'. The concepts of immersion and presence are much debated; a great number of varying definitions and interpretations exist (see for example Smith et al., 1998, Slater, 1999). Presence and immersion are sometimes used interchangeable, both denoting 'a sense', feeling or user experience; drawing parallels to movies and daydreaming. Distinctions between the two have however been made in the endeavour to clarify the issues involving the concepts. *Immersion* can be used to denote a technical property of a system. This is in line with Slater and Usoh (1994), who state that immersion is a technical property that can be achieved in various grades. In Slater's 1999 article, following a debate over contradicting notions of immersion, Slater even feels the need to change the term immersion into 'system immersion', an initiative that not many authors have taken up.

Immersion in turn may lead to a sense of *presence* for the user. Presence is defined as a state of consciousness, the sense of being there in the world that has been created by a Virtual Reality system. It should be noted that immersion is named as a prerequisite, but not sufficient, for inducement of presence. The many discussions on the concept of presence illustrate that not only technical developments and software engineering issues come into play in developing Virtual Reality systems. The mental experience VR could possibly induce in some users is intriguing. Some early VR research even makes comparisons to drug use and worries of possible dependency of users to the artificial world (as discussed by Chesher, 1994). However, these mental aspects are not the focus of this thesis. Exploration of presence as a state of consciousness is left to other researchers and focus here is on immersion in the 'system immersion' interpretation.

A necessity for immersion is that at least one sensory modality provides the illusion that virtual objects are in the same space as the user, in current VR development this is typically the visual (Smith et al., 1998). Displayed sensory information about the user's 'virtual self and body' can be matched with the user's mental model of the virtual environment and his/her own position and behaviours. According to Smith et al., the degree of immersion can be increased by the adding the number of senses that are involved. Slater and Usoh (1994) name adding additional and consistent modalities, improved and increased body tracking, richer body representations, decreased lag between body movement and resulting changes in sensory data and so on. Interaction techniques are named as possibly playing a crucial role in achieving immersion and presence. To achieve these, interaction should be 'natural' and actions should not have to be achieved in a cumbersome artificial way. What natural interaction exactly is, however, is not yet established.

Immersion is handled in the context of this study as the technological property determined by the visualization and interaction possibilities of a system. Authors like Kalawsky (1999) categorize VR systems according to immersion in desktop, semi-immersive and fully immersive systems. Following characterizations for various VR categories are used here; fully immersive systems are considered systems that surround the user with a virtual world e.g. head-mounted displays, or systems offering the user the possibility to actually be in the system (such as CAVE systems, room-sized systems which the user can stand in). Semi-immersive systems are other stereoscopic visualizations in keeping with Marsh' three key concepts of Virtual Reality. Non-immersive systems will be defined as non-stereoscopic VR applications. Desktop VR can be offered either semi-immersive with e.g. shutter glasses or non-immersive without specialized equipment. Further definitions for the various existing types of VR systems, terminology and interpretations used in this study are discussed in the following section.

2.1.6. Types of Virtual Reality systems

The frameworks of Zeltzer (1992) and Marsh et al., (1998) distinguish VR from other technologies, but VR itself encompasses a wide variety of systems. Systems that can be positioned on Milgram's Mixed Reality continuum and a number of other types of visualization systems are often categorized under the Virtual Reality header as well. Various categories used in VR related research are listed below (inspired by Belleman, 2003). The categories are not mutually exclusive, immersiveness is used to categorize various types of VR, and for many type and various interpretations exist for the terms used. Following characterizations for the listed terms will be used in this thesis.

- *Immersive VR* offers the user a stereoscopic, head tracked, as much as possible surrounding visual experience using either head-mounted displays or (multiple) projection screens, such as in the CAVE environment. Systems are deemed *semi-immersive*, when an all-around picture is not offered. The concept of immersion has varied interpretations, further exploration of the concept of immersion is offered later in this thesis.

- *Desktop VR systems*, sometimes called Window on the World systems, refer to those systems that offer a view of a virtual world via a desktop computer, either with or without stereoscopic vision. The term fish tank VR is sometimes used if head tracking is included in the system. It was introduced by Ware et al., (1993) as a term for 3D scenes on a limited size monitor with perspective projection coupled to the user's head position (Mulder and Van Liere, 2002). In contrast to other authors, fish-tank VR is here not used to refer to desktop VR applications.

- *Fish tank VR* will be defined as provided by applications that make use of equipment that enables users to actually 'reach in' and grab an object, such as virtual workbenches. The term *near-field VR* is sometimes also used for systems that allow users to interact with virtual objects within arm's reach of the user (Mulder and Van Liere, 2002; and e.g. Von Wiegand et al., 1999). This term however does not seem widely prevalent. Many examples of fish tank VR systems in this interpretation can be found in literature (see e.g. Poston and Serra, 1996; Reachin Technologies, 2004; and Mulder et al., 2003).

- *Augmented reality* is the term used for those applications where a computer generates information that is added to the user's view of the real world, through for example transparent displays (see e.g. Bajura et al., 1992; Wildermuth et al., 2001).

- *Telepresence* refers to for example operating remote controlled robots to explore real environments that would be dangerous for humans, or where it is otherwise less feasible for the operator to be present at the location of the equipment. Telepresence does not fit the definitions of VR where an artificial environment is offered (e.g. Lanier's original definition) nor does it fit the definition used here. However, VR and telepresence research communities are related and therefore this category is named here

The pictures in fig. 6 illustrate the great variety in VR systems; a number of examples of VR systems are shown, following the categories and terminology used in this thesis.

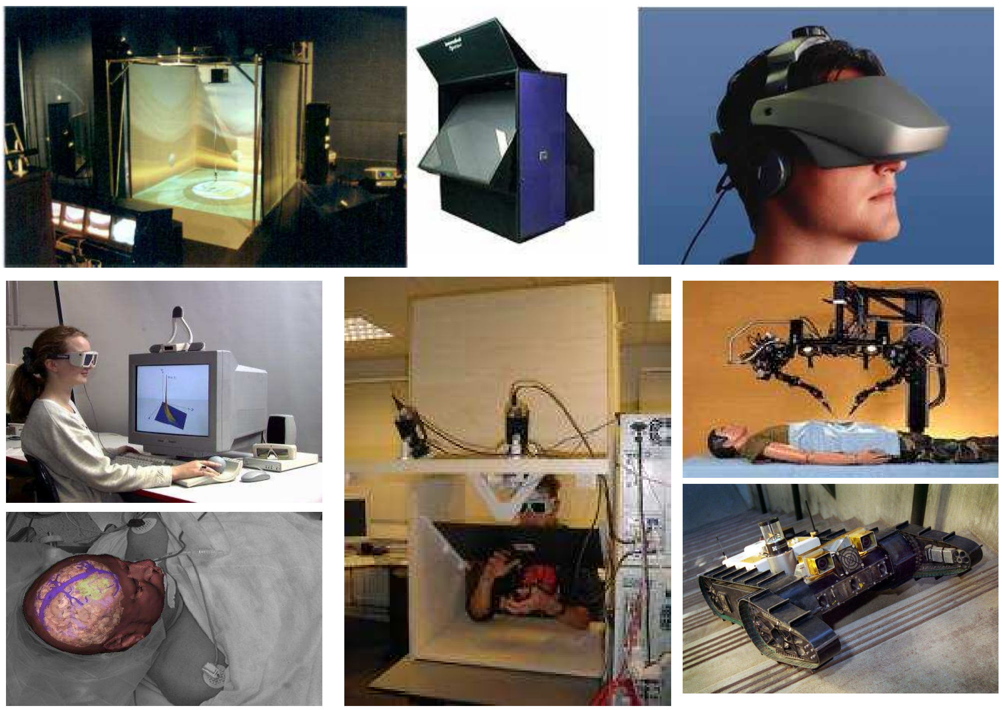


Fig. 6 Examples of VR (and related) systems. Top row, from left to right: CAVE system at SARA, The Netherlands (NCSA, 1999), example of a semi-immersive projection system (NCSA, 1999), example of a head-mounted display (VR Solutions, 2004). Second row: example of a desktop VR system (DFN, 1999), telepresence: remote surgery (SRI, 2004). Large middle image: Personal Space Station, fish-tank VR system following the terminology used in this thesis (Mulder and Van Liere, 2002). Bottom row: augmented reality projection of internal patient data on actual patient (Jolesz et al., 1997) and telerobotics to explore hazardous environments (NASA JPL, 2001).

Choosing between types of VR

Choosing between the types of VR systems that are outlined above is not a straightforward matter. Guidelines for choosing the most suitable category and immersiveness of VR systems for a particular application in specific context of use and the appropriate degree of immersion are scarce. Additionally guidance if the choice for VR in general is warranted at all is very limited as well. There is no complete framework available facilitating the choice if e.g. a (non-VR) desktop application, desktop Virtual Reality, fish tank, semi-immersive or completely immersive VR application might be the best option for a particular system, its users and their context of use and for the tasks to be supported. To develop useful and usable (VR) systems ample consideration is needed of the choice for one of these options, studying the usability criteria for an application and tasks, and analysis of the context in which systems will be used. Categorizing systems is not a goal it itself, but can offer much needed clarity in a field where a great variety of terms and interpretations of such categories are used. Focus here will not be on categorizing systems per se, but instead on usability and contextual issues that might make a certain type of (VR or non-VR) system more suitable in a particular situation. The following section will first explore the concept of usability and will go into the specifics of VR usability.

2.2. Usability

One of the main activities in the field of Human Computer Interaction is striving for useful and usable systems, with 'usability' being one of the central concepts in the field (Löwgren, 1995). Usability is the concept of satisfaction users can derive from using a system's functionality (Grudin, 1992 as quoted by Hilbert and Redmiles, 2000). A system's 'usefulness' depends on its usability and utility. Utility refers to the capability of the system functionality to support user's needs (the concept of utility can thus not be used interchangeable with the term functionality). Usability can further be defined as encompassing both ease of use and acceptability of a product (Bevan et al., 1991, Bevan 1995). Ease of use determines whether a product itself can be used. Acceptability determines whether a system will be used and in what way. Both ease of use and acceptability are specific to a particular class of users, for a particular task, in a specific context. Usability is a multidimensional concept and various, be it similar, interpretations exist. While the insight is prevalent that paying attention to the usability of a system is important, and the term is widely used, a number of different definitions and perspectives have been in existence (Löwgren, 1995).

To gain understanding of the issues surrounding usability of Virtual Reality systems, it is useful to explore the general concept of usability. This section discusses different perspectives on usability and explores the issues that come into play when usability needs to be assessed for a specific type of system in its (prospective) context of use. Issues surrounding VR usability and methods available to support development of usable and useful Virtual Reality systems are discussed as well.

2.2.1. Perspectives on usability

Löwgren (1995) describes a number of perspectives to usability as they had emerged up to 1995. The concept of usability started out with a *general perspective* where the HCI community tried to accumulate as much information about human-computer interaction as possible. Studies focusing on this perspective focus on laboratory experiments and understanding user behaviour. A more practical and system focused *usability engineering* perspective emerged to counter the limited applicability of the general perspective on usability. This viewpoint focuses on the practical applicability and operationalization of the concept of usability in guidelines and measurements. Central to this viewpoint is the separation of usability and utility as subcategories of a system's 'usefulness'. This is analogue to the way 'functional' and 'non-functional' requirements can be defined. Non-functional requirements include usability of a system. The *subjectivity* perspective takes into account that usability is subjective and depends on users and their contexts. Usability and utility cannot be separated when strictly adhering to this view. The *flexibility* viewpoint broadens the view of the usability engineering view by pointing out that a context of use is not static and constantly changes. The s*ociality* perspective takes into account that interaction does not take place in a social vacuum and is part of the field of computer-supported cooperative work.

These perspectives are not necessarily as contradictory as they might seem, but rather can be viewed as additions to the other viewpoints. Possibly one could argue that all the operationalized viewpoints should be addressed in system development. Usability and utility, for example, can be distinguished as concepts, as in the usability engineering viewpoint. This still does not necessarily contradict that usefulness of a system depends on how both utility and usability are supported (subjectivity viewpoint), E.g. Hilbert and Redmiles (2000) note that usability evaluations often address utility issues as well. Nor does it contradict that usefulness of a system also depends on the system capabilities in a dynamic context of use (flexibility viewpoint), and on the degree a system supports user cooperation if necessary (sociality viewpoint). It should be noted that in this thesis the term usability engineering is used for all activities that aim to make a system more usable and not specifically to activities only that are in line with the 'usability engineering' perspective as outlined above.

2.2.2. Defining usability for this study

The definition of usability used in this thesis is ISO standard 9241-11 (1998, as quoted in e.g. Bevan, 2001; Travis, 2003; Van Welie et al., 1999) The ISO standard describes usability as: *The extent to which a product can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*. Effectiveness in this definition refers to the accuracy and completeness with which users achieve specified goals. Efficiency is defined as the resources expended in relation to the accuracy and completeness with which users achieve goals. Satisfaction refers to the comfort and acceptability of use by end users. Claims of Abran et al., (2003) may be justified that the ISO standard 9241 as a whole in its recommendations does not tackle all aspects of security and learnability as recommended in literature and various ISO usability standards exist. However, the definition in itself provides for all aspects of usability named in literature and fits the purposes of this study.

The definition above explicitly takes into account the importance of considering the specificity of users, their goals (accomplished by executing tasks) and environment in establishing usability. A system that is usable in a laboratory set up might not be as practical in the proposed context of use. The acceptability of an implemented system does not depend only on usability of its interface; the system should fit the physical, social, organizational and technical context in which is will be used (Karat, 1997 and Bevan and Macleod, 1994). General usability guidelines can be devised, illustrated by the aspects of usability in the before section, but the specific usability demands for a system depend on the particular context of use. The use of general usability goals and guidelines for getting insight in the usability of a system is acknowledged here, while still putting emphasis on the importance of considering characteristics of users, their tasks and environment in performing a complete usability analysis for the system at hand.

The usefulness of usability guidelines does not only depend on the context of use. The type of system itself should play a role in choosing ways to evaluate the system and may cause a need for adjusting guidelines and methods. A VR system, for example, bears with it different usability concerns than a desktop application. The next section will discuss these usability considerations specific to VR systems.

2.2.3. Virtual Reality usability

Virtual Reality is sometimes considered a way to overcome limitations of human-computer interfaces and find perfect 'natural' interaction with the computer. Some authors have even thought VR to be the natural evolutionary conclusion of the progression in human-computer interaction (as discussed by Gobbetti and Scateni, 1999; Chesher, 1994). However uplifting and motivating these views might be, this perceived potential has yet to be realised; achieving usable VR systems is still considered challenging (Stanney et al., 1998).

Not only are VR interfaces generally unfamiliar to users (Bowman and Hodges, 1999), but using them also brings along new issues that do not come into play when dealing with 'traditional' 2D desktop applications. Such issues include presence, physiological effects, issues of locating and manipulating objects, or navigating in 3D worlds. Traditional usability methods do not take these factors into account (Kalawasky, 1999; Steed and Tromp, 1998; Kaur, 1998; Kaur et al., 1999; Neale and Nichols, 2001 and Willans et al., 2001).

Some specific factors that distinguish VR usability from usability concerns for e.g. regular desktop applications are outlined below as well as a number of research themes that are topical in research into human factors and VR.

VR-specific usability factors

Human factors research for VR is quite a varied field and below a number of issues in VR usability research are discussed that are specific concerns in evaluating VR applications. Each of the issues outlined has sparked various lines of research. A number of other general themes can be identified that set usability research of VR systems apart from desktop evaluation, which are discussed below and listed in fig 7. More extensive overviews are offered by e.g. JTAP (2000 project report tailored to Virtual design environments in education), Kaur, 1998; Gabbard, 1997 and Stanney et al., 1998. The overview below is not claimed to be complete, but is intended to provide insight into a number of important issues concerning VR usability engineering.

Selection of VR usability issues:

- Psychological factors
- Immersion and stereoscopy
- Navigation and orientation
- Physical effects and ergonomics
- (Multimodal) interaction

Fig. 7 Specific VR usability issues discussed

Psychological factors

The perceived potential of VR to offer the powerful experience of an alternate reality has occasionally fuelled worries that that use of VR could lead to psychological trauma, addiction to the offered alternate reality or negative changes in behaviour - just as VR can be used to positively change users in e.g. phobia treatment (see e.g. Stanney et al., 1998). Especially in the past potential harmful psychological effects and societal problems have been an issue in press and literature (for a discussion, see Chesher, 1994). However, these concerns are not fundamentally different from fears of effects of using other technologies and media. Similar discussions about e.g. aggression in children relating to television programs and 'Internet addiction' regularly flare up in the media. These issues are possibly more likely to be of importance for virtual environments offering explorable virtual worlds and virtual events. Such psychological issues are not the focus in this thesis, however it is useful to note these have been regularly brought up and could be of concern to some prospective VR users and might affect adoption of VR applications.

Immersion

Presence and immersion have been hot topics in human factors research for VR environments and related to possible psychological effects of VR use. Especially for virtual environments that offer an artificial world to be explored, such as architectural walkthroughs, the feeling of 'being there' is considered an important issue. Some research has tried to assess the influence of immersion and presence on task performance (Slater, 1999). As useful as these endeavours might be, it can be pointed out that presence might not necessarily be what matters most to users' performance. When the primary user interest is achieving insight into a 3D object, they might not necessarily be interested in experiencing presence in the virtual environment; offering an experience that mimics the feeling of being present might not be important. Exploring and 'experiencing' a virtual environment around the object of interest that offers no additional information on that object might be unnecessary and possibly even detrimental to user performance.

It can be argued that usefulness of a system should be the focus of development and not immersion per se. Immersion is possibly not necessary for every application and the task, user and context of use it aims to support. It is yet unclear which kind of application and context of use warrant the choice for immersive VR and e.g. stereoscopic vision. Should perhaps an application that does not aim at presence be implemented in an immersive VR system at all? Similarly, the question can be asked for what tasks and in what context stereoscopy is really helpful and important. No definite answers are available here.

Navigation and orientation in VR environments
When an explorable VR environment is offered, navigating of this virtual world can pose problems as well. Finding your way around in virtual environments that closely mimic their real counterparts is still more difficult than navigation in real life (Ruddle et al., 2004). Users can potentially become disorientated and effectively get lost in the virtual world. Locations of items of interest could be difficult to find or remember for the user. McNeill et al., (2002) sum up a number of such issues related to navigating virtual environments identified in literature:

- Getting lost or becoming disoriented in the environment frequently occurs when the speed of navigation is too fast or the direction of movement is not as expected. Navigation speed needs to be appropriate for the size of the scene and the tasks that have to be carried out.
- Users can become disoriented if movement is not confined to a level plane
- Lack of identifiable landmarks in the VE aids to disorientation- objects in the virtual environment that can be used as points for orientation and cues in wayfinding
- Actions that are impossible in the real world such as flying, speeding up navigation, moving through objects and boundaries such as walls and 'teleporting' to other locations can be useful, but can also add to disorientation.

Various techniques to counter disorientation in VE's have been explored. Constraining and guiding navigation might produce more effective navigation and orientation. Guidance can be provided by e.g. landmarks, maps (Haik et al., 2002), virtual footprints (Mourouzis et al., 2004) and 3D worlds-in-miniature virtual models used within the large virtual counterpart (Pausch and Burnette, 1995). However, orientation and navigation are still issues to deal with in designing a VR system.

Interaction with Virtual Reality systems
Virtual Reality systems often use a variety of interaction modalities. Besides pointing with a 'space mouse', body-position tracking, voice recognition and haptic interaction have been used. Various modalities can be combined and each of these possible modalities offers various options in ways of interaction, for each modality various hardware devices and software configurations are possible. The general consensus is that there is no one best interaction modality; each has its merits as well as disadvantages. There is however no complete overview of which modality and device fits a situation best.

An example of endeavours towards such an overview is the research that is carried out on combining speech recognition with other modalities. McNeill et al., (2002) for instance are confident that 'spoken dialog systems for navigation can offer real advantages over traditional mouse and keyboard interaction' and state that while mouse and keyboard interaction may be optimal for interaction with virtual objects they can be clumsy for navigational purposes (McNeill et al., 2002). This while in an experiment by Stedmon et al., (2004) participants felt that speech was good for discrete tasks (like changing an object's colour), but performed less well in general navigation.

It has to be kept in mind that better or worse performance with a particular modality might also be the consequence of using a particular configuration of that modality. An example of examining performance with various configurations of a modality are e.g. the experiments on use which kind of menus are most convenient to use in virtual environments (e.g. Serra and Waterworth, 1996). Performance with a certain modality cannot be viewed separately from the context of use of a particular application. This can be illustrated with the use of speech recognition that might be convenient for a user in a private room, but socially unacceptable when speech would disturb others in an office. Performance on a modality can also not be separated from ergonomics; using a particular menu might result in more required movements by the users or uncomfortable postures, and depend on human haptic preferences (Lindeman et al., 1999).

Ergonomics of Virtual Reality systems

Especially for VR systems, not only the graphic look, but also the actual physical feel and ergonomics of an interface are important (Hinckley, 1996). Using immersive VR systems might be more of a whole body experience, especially standing 'in' a system such as in a CAVE, rather than sitting in front of a desktop monitor. Design and evaluation of desktop applications usually do not seem to address ergonomic hardware issues, such as possible repetitive strain injury that might result from using a mouse. In contrast, developing a VR system is usually focused at delivering a whole system, including hard- and software. This is especially the case when such specialized systems such as flight simulators including a mock cockpit or operating trainers using faux medical instruments are developed.

Specialized VR input devices are constantly being developed; there are yet no standards for a 'space mouse' as there are for a desktop mouse. VR interaction devices vary wildly and can range from desktop mouse-like devices with an added sensitivity on a third dimension, to 3D objects such as cubes (Mulder and Van Liere, 2002) or even doll heads (Hinckley, 1996). E.g. Mulder and Van Liere (2002) offer possibilities to develop unique interaction devices from scratch using 'any' object available (whether a pen, a wooden block, a little ball or something else) and sticking on highly contrasting stickers to make a pattern that the system can visually recognize the item with. Such new possibilities for 3D input devices, offer their own challenges, and raises questions on when they are of real advantage (Hinckley, 1996), what devices would be appropriate for a specific application and how devices can be improved.

Not only ergonomic qualities of input devices, but also the consequences of VR use on overall physical well-being is of concern; nausea, dizziness and eye-strain are often reported while and after VR use. Various container terms such as ‘Virtual Reality Induced Symptoms and Effects’ (VRISE) (Cobb et al., 1999), are used to refer to these physical discomforts, which can be quite prominent and detrimental to overall usability of VR systems. Effects of using immersive VR systems on vision have been researched as well, especially for head-mounted displays. Immediate harm has yet to be (dis)proven, but suboptimal settings can result in discomfort and eye fatigue (see e.g. Stanney et al., 1998).

Other physical health concerns also include problems concerning users wearing extensive headgear that blocks out the real world. Tripping over 'invisible' wires or getting tangled in them might not only be a possible nuisance to the user, but also a health hazard (and a lawsuit waiting to happen) as are the blocking out of any auditory signals from 'outside'. Bowman et al., (2002) even offer advice for VR system evaluators to shield experiment participants from possible harm. It can be noted here that such authors as Stanney et al., (1998), identify shutting out the real world as a factor in achieving immersion, complete immersion then, could be argued to pose serious practical safety risks, especially when all modalities are blocked out. Users will probably not accept serious physical discomfort if they have to use a system repeatedly and should not be put in any physical danger. Attention should be given to such, and less dramatic, ergonomic issues in designing and evaluating a VR system.

2.2.4. Evaluating usability of Virtual Reality systems

To address the issues outlined above a number of challenges have to be met. First of all evaluating a VR system with users can bring with it considerable challenges in e.g. finding enough participants and avoiding that these participants get simulator sickness. Apart from user evaluation of system, quite a number of methodologies have been applied in VR evaluation. New methodologies have however been developed as well, since many available usability evaluation methods do not take into account specific issues of VR evaluation, such as the 3D nature of interfaces and navigation through a virtual world. These methods are mostly based on expert walkthroughs and developing new heuristics for evaluation.

A number of usability evaluation methods that are specifically aimed at VR systems are discussed in this section. To illustrate the practical challenges a evaluator can encounter, the following section will however first list a number of issues that can come into play while evaluating a VR system with users.

Challenges in Virtual Reality usability evaluation

Apart from the before mentioned ergonomic issues, evaluating a VR system can pose difficulties. Bowman et al., (2002) offer a rather comprehensive overview of challenges that can arise specifically during user evaluation of VE's. They name a number of user issues, evaluator issues, physical environment issues (with which they refer to the physical properties of the VR system, not its context), and evaluation type issues.

First of all, it is often not possible to find large numbers of representative users in evaluation of a VR system. Not many people are experienced users of Virtual Reality applications. A certain 'wow' factor might dazzle first time users of virtual environments, which might result in an overly positive evaluation. A contrast can be made here between VR entertainment systems that actually aim to achieve this wow effect, and systems that aim to support a work task. Especially evaluation of the latter with inexperienced users can miss usability problems that would be apparent in day-to-day use of a system. An added problem is that often, especially in case of larger physical prototypes, prospective users also have to be brought into the lab, requiring considerable effort and time investment from participants in the evaluation.

For the evaluator it is difficult to find out exactly what the user is experiencing. The user is not typically sitting behind a desk, but uses equipment and walks around, and views stereoscopic images from a certain viewpoint. What the user sees exactly can be difficult to capture, especially when using displays such as head mounted displays that cannot be seen directly by others. Evaluator interventions in an evaluation might also disturb the experience of the user and affect evaluation outcomes, especially when issues such as presence are studied.

During evaluation physical comfort of participants should be monitored as well. Due to the possibility of 'Virtual Reality Induced Symptoms and Effects' (VRISE) symptoms like nausea and fatigue, the strain put on them might become unacceptable. Accidents like getting tangled in equipment wires or bumping into display screens are not unlikely in many VR systems either, especially where a certain suspension of disbelief is achieved. It is unacceptable for participants to be in physical danger while using a system,; accidents and serious discomfort should be avoided.

An additional often-occurring problem with evaluating Virtual Reality systems is their prototype stage. System software and hardware are often still under serious development and unstable, which can be a serious nuisance and can interrupt evaluation sessions. The goal of an evaluation session might be to identify such problems as well, but VR systems often are in such a stage of development that a complete evaluation becomes difficult. Since VR systems are not seldomly endeavours used to develop complex techniques instead of proposed solutions to existing problems of potential users, it can often also be unclear what requirements the system under development should fulfil and how it should be evaluated.

For a more comprehensive overview of issues in VR evaluation the interested reader is referred to Bowman et al., (2002), who provide a wealth of examples of challenges that may arise during VE evaluation, especially during user tests.

Whether or not evaluation takes place with or without users; an appropriate evaluation method has to be selected; at this time a considerable challenge as well. For evaluations of VR systems a number of methodologies and tools exist, the following section will elaborate on these available methodologies.

Virtual Reality usability evaluation methods

Traditional usability evaluation methods are tried and tested for desktop applications, but Virtual Reality applications are often radically different from desktop applications and the usual methods may not be applicable. Heuristic evaluation for example, is deemed challenging due to the current lack of well-formed guidelines and heuristics for virtual environments design and evaluation (Bowman et al., 2002). Commonly used heuristics like e.g. Nielsen's heuristics do not address such issues as 3D interaction and navigation in a virtual environment. Many usability evaluation methods and data collection techniques, including heuristic evaluation, used in usability evaluation for all kinds of (computer) systems have however been applied to VR evaluation. Bowman et al., (2002), give an overview of various of usability evaluation techniques that have been applied in VR research. Fig.8 shows their categorization.

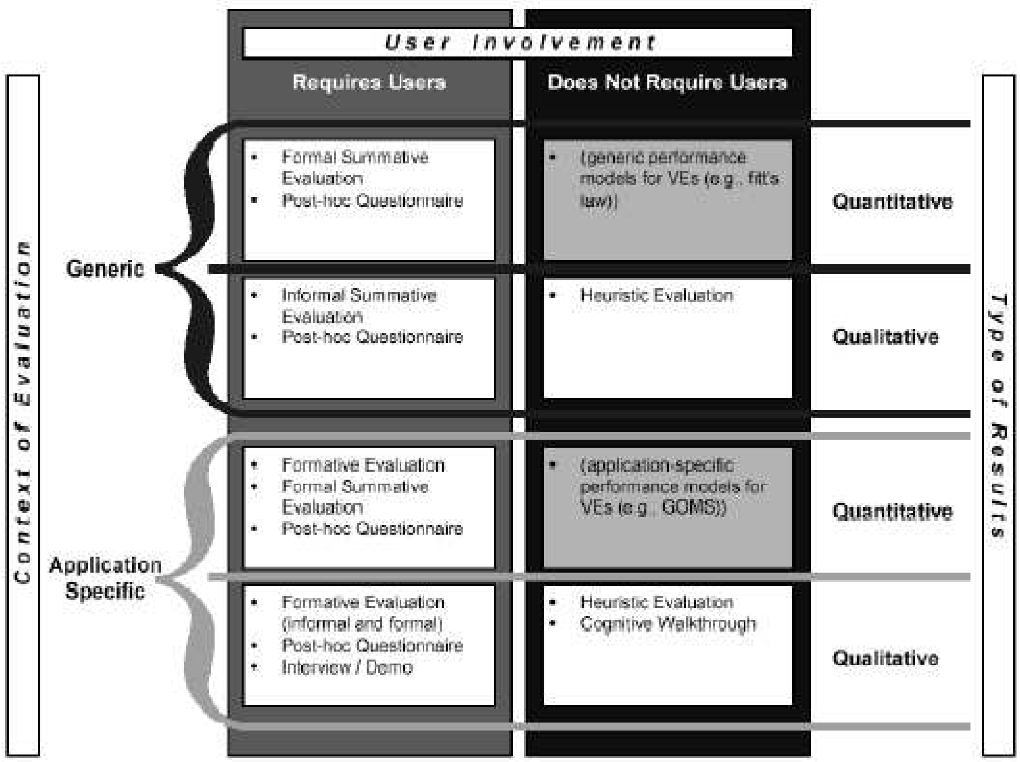


Fig. 8 Methods for usability evaluation that can be applied for evaluation of VR systems (Bowman et al. , 2002).

Bowman et al., categorize the usability evaluation methods on three criteria: type of results, user involvement and context of evaluation. The grey shaded sections in the table denote these combinations of the criteria that yet have to be applied for VR systems. Bowman et al., describe both general methods that have been applied in VR system evaluations and a number of methodologies specifically aimed at VR applications. They show that most of existing usability evaluation methods have been applied in VR related studies, but both generic and application specific performance models have not yet been developed.

A number of specific methods aimed at VR are discussed in this section, the following table 1 gives an overview.

Authors	**Method or tool**	**Short characterization**
Willans and Harrison (2001)	Marigold technical tool	Tool aiding quick implementation and evaluation of VE interaction techniques.
Bowman et al. (1999, 2001)	Testbed evaluation	Method for testbed evaluation of VE interaction techniques.
Neale and Nichols (2001)	Theme based content analysis	VE evaluation method based on categorizing usability problems encountered by users. No heuristics offered. Based on content analysis from social research.
Gabbard et al. (1997, 1999)	Sequential heuristic and formative VE usability evaluation	VE evaluation method alternating heuristic and formative evaluation with users performing representative tasks. Informal set of 195 guidelines is offered. Based on Norman's theory of action.
Sutcliffe and Kaur (2000) Kaur et al. (1999)	Walkthrough evaluation	User task-based evaluation with questionnaires based on 64 Generic Design Properties (GDP's) a usable system should embody. Based on Norman's theory of action.
COVEN (1997) Steed and Tromp (1998)	Heuristic evaluation and cognitive walkthrough	Heuristic evaluation and cognitive walkthrough of VE's using adaptation of Nielsen's 10 usability heuristics.
Kalawsky et al. (1999)	VRUSE	Questionnaire for usability evaluation of VE's using 10 VR usability factors operationalised in ±100 questions. Adaptation of MUSiC framework for usability evaluation.

Table 1: Various available usability evaluation methods for VR systems.

Below first the two methods of Willans et al., and Bowman et al., specifically aimed at evaluating VR interaction techniques are highlighted. Available methods that use walkthroughs, e.g. those based on Norman's theory of action such as Gabbard et al., and Stufcliffe and Kaur are addressed as well, along with methods by other authors that offer specialized heuristics or questionnaires. The methods are introduced shortly and some (main characteristics and (dis)advantages of the particular method are highlighted.

Evaluation of VE interaction techniques

A complete analysis of a virtual environment also takes into account the interaction techniques that provide the mapping between the user and the virtual environment; the series of user actions, including those on input devices that lead to a certain result in the virtual environment. Willans and Harrison (2001) and Bowman et al., (1999, 2001) focus on the evaluation of these interaction techniques. Willans and Harrison present Marigold; a technical aid generating code for different interaction techniques. They try to offer easy experimentation with different device-technique configurations so both capability and usability of interaction techniques can be evaluated. A technique to evaluate the interaction techniques or VR system itself is however not offered. Bowman et al., (1999) do aim to offer such a methodology for design, evaluation and application of interaction techniques.

Bowman et al., claim that interaction techniques should be well adjusted to basic tasks so they can be applied to a wide range of complex tasks. They identify three task categories of VE interactions that could be used as the basic building blocks for most interactions in virtual environments; selection, manipulation (positioning and orienting objects) and viewpoint motion control (or travel) and a forth compound category of 'system control', tasks that are a combination of election/manipulation, such as saving a current location or e.g. changing an objects colour by selecting the object, selecting a colour and giving a 'change colour' command. In their methodology 'test beds' are devised to compare performance with various interaction techniques on these tasks. These test beds are scenarios that involve all important aspects of a task and that can be used to evaluate each component of a technique and provide multiple performance measures.

Testbed evaluation could be useful in developing the guidelines on what interaction techniques are appropriate in particular situations. Willans and Harrison's Marigold may be a very useful tool to speed up the process in which interaction techniques can be compared by providing support for implementing alternative interaction techniques. Even though evaluating interaction techniques is necessary as a part of a usability evaluation, developing interaction techniques and these described methods are not focused on in this thesis. For a practical and complete evaluation of a VR system other methodologies that go beyond evaluation of interaction techniques are necessary; such methods are outlined below.

Theme-based content analysis

Neale and Nichols (2001) present theme based content analysis for VR systems; an adaptation of content analytic methods from social research applied to system evaluation. The method is based on evaluating a system with users performing tasks in the virtual environment and then categorizing their remarks and problems that they experience. Data may be collected using any method that yields verbal data of users. Multiple evaluators individually group the raw data into themes and then come together to determine a classification of higher order themes. The aim of theme-based content analysis is to swiftly identify usability problems and help in iterative development of systems.

The method seems very similar to cognitive walkthrough evaluation methods, why the method would be specific to Virtual Reality seems somewhat unclear. Theme based content analysis even seems to offer no specific guidance for VR systems at all. Neale and Nichols doubt the possibility of a complete and useful set of usability guidelines for Virtual Reality. They do determine four factors in the user experience of Virtual Reality (system design, virtual environment design, Virtual Reality implementation and individual user characteristics), but do not incorporate those in their theme-based method. Even if the flexibility of the method could be useful in a fast developing field as Virtual Reality, it would perhaps be useful to offer evaluators more guidance on what kind of problems to look for in evaluating a VR system. In the following sections a number of methods are discussed that offer such guidelines.

Methods offering heuristics and guidelines

The Following methods offer heuristics that can be used in VR system evaluation, or present questionnaires from which guidelines can be derived.

Gabbard et al.'s sequential evaluation, including heuristics

In contrast to Neale and Nichols, Gabbard et al., (1999) do offer more guidance specific to VR usability. An informal set of 195 usability guidelines is used for their heuristic evaluation method (for an overview, see Gabbard, 1997). After heuristic evaluation, they recommend to explore potential usability problems by placing representative users in task-based scenarios to assess the ability of the interface to support user exploration, learning and task performance (Bowman et al., 2002). Gabbard et al.'s organize the usability guidelines in a taxonomy based on Norman's theory of action. The four parts of the taxonomy are mapped on Norman's stages of activity and associated interdependencies in interaction between human and machine.

Practically, the guidelines offered do seem to address a very wide array of aspects of interaction with virtual environments. Some of the subcategories of the four taxonomy areas seem largely to be a result of the availability of guidelines from studies that can be bundled together in a category, not from a further theoretically motivated subdivision. Gabbard et al., state the taxonomy still should be expanded, the number of guidelines (195) however seems already somewhat overwhelming.

Even if only guidelines relevant to the application at hand should be implemented, all taxonomy areas still have to be examined by evaluators to determine which guidelines are applicable to the system at hand and are also not outdated due to technological developments. Some guidelines might be too device specific to provide room for future technological developments; e.g. 'data gloves' and 'space mice' and explicitly named as categories. Even of this specificity might help for users of the taxonomy to understand and apply the guidelines, possibly a more general taxonomy could be developed, applicable to a wide range of past, present and future virtual environments. The guidelines seem however , regardless of their shortcomings, to be useful in heuristically evaluating a VR system and guiding practical development.

Sutcliffe and Kaur’s walkthrough evaluation and heuristics

Sutcliffe and Kaur (2000) introduce a specialized walkthrough method for evaluating the usability of Virtual Reality user interfaces, similar to Hix and Gabbard’s approach. It is primarily aimed at reviewing desktop VR systems, but some of the recommendations may be applicable to immersive VR systems as well. The evaluation method consists of performing representative tasks within the system and filling out checklists of questions. Usability problems and missing features are then listed and prioritised.

Sutcliffe and Kaur's walkthrough analysis method uses three checklists for three categories of tasks. These categories are, like in Hix and Gabbard’s approach, extended from Norman's theory of action model. The resulting categorization of guidelines however is very different from their approach. Instead of presenting models for input, output, the virtual model and users and tasks, Sutcliffe and Kaur categorize guidelines in three models according to tasks in the virtual environment:

- The first model covers goal-oriented task action.
- the second covers exploration and navigation in virtual worlds.
- the third covers interaction in response to system initiative.

The questions of the three corresponding checklists’ are linked to forty-six Generic Design Properties (GDP's). These are properties that an usable system should embody to support successful interaction. They express abstract requirements and are e.g. affordances, knowledge sources and indicate possible causes for interaction problems. GDP's serve as starting points for concrete design guidelines (Sutcliffe and Kaur, 2000). A checklist question like ‘Are the appropriate objects or parts of the environment visible?’ is for example coupled to the more abstract GDP of ‘clear environment structure’.

Like Gabbard et al.’s large number of heuristics, the long list of Sutcliffe and Kaur’s GDP’s seems useful as concrete guidance to evaluators, but seem somewhat overly comprehensive for efficient evaluation. Their more abstract nature does however make them more generally applicable, and the three offered checklists (which curiously name only little more than 20 of the originally 64 GDP’s) seem specific and concise enough for a meaningful and practical evaluation offering guidance for improvement.

Steed and Tromp’s walkthrough evaluation and heuristics

Specialized heuristics for VR systems are also offered by Steed and Tromp, who describe their experiences in evaluating VE’s using heuristic evaluation and cognitive walkthrough (Steed and Tromp, 1998; COVEN , 1997).

Steed and Tromp have adjusted general usability heuristics by Nielsen (1994, see fig. 9), for the early stages of evaluation of VR prototypes. They applied their heuristics in the COVEN project in a collaborative virtual environment (1997). For this specific project, Steed and Tromp have adjusted the following of Nielsen's categories:

- **User control and freedom,** added is the interpretation that flexibility is needed in navigating the virtual environment.
- **Flexibility and efficiency of use,** for the early COVEN-project prototype Steed and Tromp stated that focus should be on efficiency of use rather than flexibility (no need for tailorable interaction, accelerators definition at this stage of the prototype).
- **Aesthetic and minimalist design,** aesthetics are important in virtual environments, however the minimalist design needs a different focus: opportunistic exploration of a virtual environment may be of importance, thus excessive 'minimizing' would impair the experience of that environment.
- **Awareness of other participants,** is added for collaborative environments. The focus is more on virtual environments where various participants 'meet' in a virtual world and not on systems that are used by several users who are physically in the same room or . Since the focus in this thesis is not on such environments, this aspect is not further explored here.

Visibility of system status: provide appropriate feedback within reasonable time.
Match between system and the real world: speak the users' language and follow familiar conventions
User control and freedom: users need to able to exit unwanted states , support undo and redo.
Consistency and standards: be consistent in design and used terms, follow platform conventions.
Prevent errors: no errors is better than explaining errors
Recognition rather than recall: make objects, actions, and options visible. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
Flexibility and efficiency of use: cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
Aesthetic and minimalist design: every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
Help users recognize, diagnose, and recover from errors: error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
Provide help and documentation

Fig. 9 Nielsen's 10 usability heuristics (1994)

Steed and Tromp do show that Nielsen's heuristics could be used in VE evaluation, but that adjustments are necessary. However, their adjustments are rather specific to the COVEN project and the prototype stage of the evaluated system, which makes wider application of their specific list seem rather difficult.

Kalawsky et al's VRUSE

Kalawsky et al. present the VRUSE method for evaluating VR systems. VRUSE is an adaptation of the MUSiC framework for usability evaluation. The tools that MUSiC offered were found to be suitable for 2D text and point-and-click desktop applications, but not applicable for a VR system (Kalawsky, 1999). VRUSE itself is a ten-part questionnaire in which each part addresses a key usability factor in an interface. For each of the ten goals a number of questions can be asked to the user after using the system that is being evaluated; in total a hundred five point Likert-scale questions are presented in Kalawsky et al., (1999).

Ten VR usability factors are specified, which represent the parts of the evaluation questionnaire:

- **Functionality:** The interface should be able to provide the level of functionality (and control) the user expects in order to complete a task.
- **User input:** The user should be able to interact with and control the virtual environment in a natural manner.
- **System output (display):** Information displayed to the user should be understood, unambiguous and necessary to complete the task.
- **User Guidance and help :**The user should be able to request help via online assistance.
- **Consistency:** The operation of a VR system should be consistent with the user's understanding and convention.
- **Flexibility:** The VR system should not constrain the user who should be able to interact with the system in a flexible manner.
- **Simulation fidelity:** In order to be useful a VR system needs an underlying model or simulation to control the virtual environment.
- **Error correction/handling and robustness:** All computer systems should provide error correction and recovery before a permanent change is made.
- **Sense of immersion/presence:** A VR system should allow a user to feel part of (or immersed in) a virtual environment.
- **Overall system usability:** Overall a VR system should be intuitive and easy to use.

The factors display a certain overlay with the standard usability categories of Nielsen, but certainly are not the same categories. Interesting to see is that ‘functionality’ is explicitly included; Kalawsky does no adhere to the common utility/usability dichotomy. However, this is certainly not unusual in usability evaluation, where utility is often addressed as well as usability. It should be kept in mind that the questions are primarily geared at evaluation with users and ‘grading’ or comparing interfaces, not at specifically offering practical guidance on how to improve a VR interface.

Comparison of VR system usability evaluation methods

The methods outline above, even if they offer specific guidance for virtual environments, are not fundamentally different from general usability evaluation methods for desktop systems. The techniques mostly revolve around a walkthrough of the system by users using representative tasks, coupled with lists of heuristics.

The offered guidelines and methods seem to stem mostly from practical experience with developing VE’s. All authors build on other methodologies, but only Sutcliffe et al., and Gabbard et al., offer a theoretical base by referring to Norman’s theory of action. Interesting is how their resulting categorizations of VE usability guidelines are very different, while seemingly neither one is ‘obviously better’, they just offer a different focus. Sutcliffe focuses on the cycle of navigating an environment, finding objects and performing actions with them. While Gabbard et al., include input and output mechanisms as well as the virtual model and users and their tasks.

When comparing all methods discussed, this difference in focus on different parts of a VR system becomes apparent as well. While e.g. Gabbard also addresses input devices themselves, Steed and Tromp focus more on evaluating the virtual world itself. Kalawsky for example also addresses simulator sickness and user’s appreciation of the used display devices. Aimed at VE’s gaining presence in a virtual environment is in Kalawsky’s questionnaire a principal topic, while Sutcliffe et al., do not even mention presence in their checklists.

The methods on offer could thus potentially be used complementary, but are also suitable at different stages of system development. The two evaluation methods for interaction methods that have been discussed first in this section are for example useful to provide the empirical backing to heuristics that can be used in further research and development of a system, while e.g. Kalawsky's questionnaire could be used to test user's acceptance of a more finalized system. The approaches of Steed and Tromp, Sutcliffe et al., and Gabbard et al., could probably be of most use during earlier stages of development of a VR system. For any project a well thought-out choice for appropriate methods has to be made and for VR system evaluation the available methods probably need to be adjusted to the system at hand and questions the evaluators have.

It needs to be kept in mind that all of these methods are focused on evaluating a VR system after the choice has been made to develop a virtual environment, not on determining if a VR system is the right choice for the application. Gabbard (1997) suggests that task completion time, task error rate and task learning time could be useful measures in making this determination for specific tasks. They could perhaps be used in this determination. Methods to help make this decision are however not yet available.

Like most usability evaluation methods, the specific methods outlined above focus on evaluating a system in isolation, outside of the prospect context of use and irrespective of whether the VR system would fit within this context. Usability of VR systems is still often approached as an attribute of a system on its own and not as a concept that is to be approached within a context of use. This is understandable as VR systems still are often research endeavours not necessarily directly meant for application in the real world and VR usability in general is often challenging. For immersive virtual environments evaluators even have to deal with two contexts: the explorable virtual world itself can become a context in which tasks and interactions are performed and the user's context outside the virtual environment, e.g. the workplace. Growing attention for the 'outside' context is however illustrated by growing research into e.g. collaborative VE's where interaction between users and not only interaction with a system is addressed. When VE's are to be applied more in the real world outside research laboratories, more holistic attention needs to be given to both relatively context-independent usability issues and to the tasks, users and environment a system will be in. The following section will focus on such contextual requirements.

2.3. Context analysis of requirements and constraints

In developing Virtual Reality systems focus is rarely on factors of the context of use and environments outside the research laboratory. Focus is still mostly on interface usability issues and the created virtual world, instead of the real world outside the system. Requirements analysis during system development needs to take into account the constraints, to which successful implementation of a system is bound. The section below will first outline what the concepts context of use and requirements entail in this thesis. Various methods for requirements engineering are discussed, of which some are more explicitly aimed at studying the proposed context of use of a system. Ethnographic approaches to requirements engineering are focused on.

2.3.1. Context of use

Considering context of use during the development process is crucial in achieving useful and usable systems. A system and its interaction possibilities can be very applicable in one setting and completely useless and unusable in the next. For example, a medical decision support system can offer adequate and useful advice on patient treatment after some minutes of filling in screens and computation, but is of no use in a context where a user desperately needs those minutes to save a critical patient in an emergency ward.

Context of use, like the concept of usability, seems to have various, similar definitions. Preece et al., (2000) define context of use as the social, organizational and physical environment in which a system will be used. Bevan et al., (1991) define context as the user, task and physical and social environment. They include users but do not explicitly name technical context, which is considered in this thesis. The ISO standard 9241-11 (1998) defines context of use as the users, goals, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used. For the purposes of this thesis, the term context of use will refer *to users and their goals and tasks, as well as the technical, social, organizational and physical environment of users that could potentially affect system use*. The following sections address these concepts and the issues in defining system requirements and constraints resulting from a context of use.

Users

User characteristics can affect system use as well. A system that is usable for one user, might not be for the next. Often used dimensions to define users are system experience, domain knowledge and experience with computers (Nielsen, 1993). Not only the amount of user's experience and knowledge, but more emotional attitudes like computer anxiety can affect the perceived ease of use of a system (Hackbarth et al., 2003). In devising a system user characteristics have to be considered to develop a system that is acceptable for the prospect users and will actually be perceived by them as useful and usable.

It needs to be considered that system users can encompass a wide group of people and that not only end -users who directly interact with a system should be considered. The concept of 'stakeholders' is used to refer to 'people or organizations who will be affected by the system and who have a direct or indirect influence on the system requirements' (Kotonya and Sommerville, 1998 as quoted in Preece et al., 2002). Stakeholders include e.g. the development team itself, managers, direct users and users of system output. Subsequently under this definition, the group of stakeholders can be quite large and not necessarily all stakeholders have to be involved in the requirements engineering process, but it is useful to consider the wider system context in achieving a useful and usable system (Preece et al., 2002).

Sharp et al., (1999) have proposed a system to identify all relevant stakeholders of a system. They propose identifying baseline stakeholders, their clients (those who the baseline stakeholder delivers information to), suppliers (those who deliver supporting tasks and information to the baseline stakeholder), and others who interact with the baseline stakeholders: the satellite stakeholders. They distinguish four different baseline stakeholder groups: users (those who use the system directly and those that use products of the system), developers of the system, legislators (ranging from governmental bodies, unions, auditors to own organizational bodies) and decision makers (managers, financial controllers, etc.). Their approach focuses on interaction between stakeholders, rather than on stakeholder-systems relations. Identifying stakeholders is important, but this approach may have the risk of not knowing when to stop. Eason (1983, as quoted in Preece et al., 2002) offer a clearer structure for defining and categorizing users and stakeholders. Eason distinguishes three categories of users: primary, secondary and tertiary. Primary users are the frequent hands-on users of a system. Secondary users use a system occasionally or use the system through an intermediary. Tertiary users are those affected by the introduction of the system or who will influence its purchase. The three types of users can together be defined as stakeholders, while the opposite does not hold.

VR research is usually concerned with direct users of a virtual environment, this might be the consequence of the focus on experiencing a virtual environment in e.g. architectural walkthroughs instead of using information coming out of using a VE. For the purposes of the study reported in this thesis, the main concern is with the direct, primary users, but other stakeholders are considered.

Environment

The concept of 'environment' as part of context of use refers to the technical, social, organizational and physical environment of users that could potentially affect system use.
First of all the technological environment of a system needs to be considered; what tools are in use now that it might replace, what tools have to be used together with the system, with which systems it will need to be compatible and share files with. The social and organizational constraints for a system need to be taken into consideration. Are there any ethical considerations for the system? Will implementing a system change working relations in a department? Does the system fit the workflow in an organization, what could be improved in the department, or what cannot be changed? The physical environment can provide for specific requirements and design considerations as well; is there enough space for the system, what are the light and noise levels of the room a system could be placed in. A stand-alone system using voice recognition might be very usable in a quiet dedicated space, but might not be the right choice for a hectic and noisy open office. Such methods as CommonKADS for knowledge systems (Schreiber et al., 1999) provide some insight on aspects of in particular the organizational context of use that might affect a system's design and development (even if the method focuses on knowledge systems in particular). What the important aspects of a specific system's environment for a system are, has to be studied however in the actual context of use. It might be too easy to overlook important considerations when not actually visiting the prospect context of use. This is especially important when developing a system that cannot be considered a 'standard' desktop application and relatively unknown issues might come into play.

2.3.2. Requirements engineering

Requirements engineering and involving users in this process is a vital part of the system development process (Preece et al, 2002). A wide choice of methods is available for requirements engineering and for gathering and structuring the data used in this process. This section describes some of these methods. It elaborates on how analysis of context-of-use fits into requirements engineering. After a discussion of involving users in the requirements engineering process, this section elaborates on methods that exist for analysis of the prospect context of use. No specific requirements engineering methods for VR systems are available and using existing methods seems reasonable.

'Requirements'

When finding out what the requirements for a particular system are, for a particular use by particular users in a specific context of use, it is useful to be as particular about what a 'requirement' is as well. A requirement can be defined as 'a statement about an intended product that specifies what it should do or how it should perform' (Preece et al., 2002). Functional requirements indicate what the system should do, while non-functional requirements state the constraints that exist for the system and its development. Non-functional requirements include data requirements, context-of-use (physical, social, organizational, technical), user (such as user skill and knowledge) and usability (goals and measures) requirements. Note that user requirements are here explicitly named to emphasize the importance of considering user characteristics, but the concept of context of use in this thesis entails the system environment, users and their tasks. The term *requirements engineering* is here preferred and used as a container term for all activities leading to a greater insight in user requirements.

2.3.3. Requirements engineering methods

A user-centred approach to requirements engineering is generally advocated. Quite a number of models, methodologies and approaches have been developed for trying to adequately find out what users need and for involving them in system development. For example Dix et al., (1998) describe a number of user requirements modelling approaches that are concerned with user needs, such as socio-technical models, soft system methodology and participatory and iterative design. Added to these methods, task analysis and cognitive models are described as possible tools in requirements engineering.

A wide choice in methods to gather data for requirements engineering is available. Questionnaires, interviews, focus groups and workshops, observation and studying documentation are all widely used tools for gaining more information on user requirements. When choosing between such techniques for data gathering in the requirements activity, a number of factors need to be considered: the nature of the technique, knowledge required of the analyst, the nature of the task to be studied, the availability of stakeholders and other resources and the kind of information that is needed. Data gathering techniques differ in the amount of time they take and level of detail and risk associated with the findings and the knowledge the analyst must have about basic cognitive processes (Preece et al, 2002). A large number of authors (e.g. Preece et al, 2002, Dix et al, 1998, Robson, 2002, Swanborn, 1997) have extensively described the pros and cons of various data gathering techniques involving human subjects, most arguments are well-known and will not be repeated here.

Focus here is on gaining insight into an actual context of use. Actually observing users during their work on the tasks a system should support in that setting seems to offer the most direct way to find out more about users, their tasks and context. Here the focus will be on ethnographic approaches to collecting contextual data and how these might be beneficiary in VR development for a day-to-day context.

Ethnography in system development

The importance of observing users during their work in their work environment for finding out about user requirements is emphasized in literature and some form of direct observation is often viewed as essential (e.g. Dix et al, 1998, Preece et al, 2002, Fouskas et al., 2002). According to Hackos and Redish, (1998) most can be learnt about users' needs and wishes by actually being with users; watching them work, observing in their physical, social and cultural environment, asking questions in that environment and listening to the users themselves as they explain what they do. Observation yields direct information about user behaviour, instead of getting, potentially inaccurate, information by using for example interviews or questionnaires.

Observation in test labs, focus groups, and surveys can be useful in finding about more about users, but they are no substitutes for actually observing, listening to and talking with users in their own work environments. Ethnographic observation is a technique that focuses on naturalistic observation; studying situations as they occur in their natural setting. It has traditionally been used in anthropological studies. An ethnographic approach to studying user requirements involves 'getting out there' in the prospect context of use of a system; observing users e.g. in the workplace. It assumes that for developing technology that supports work, the work needs to understood as a social phenomenon (Löwgren, 1995).

Even if anthropologists do carry out studies in system development, see e.g. Bell et al., (2001), but there are some differences with 'traditional' ethnography. The difference of traditional anthropology ethnography with applying ethnographic methods in requirements engineering lies mainly in the duration of the observation. In anthropological work researchers may live with their subjects for months or years, while observation in engineering may take only hours. Another difference is that in the social sciences usually ethnographic observation takes a directly interpretative form where actions observed are categorized and codified in action types that have a relation with a particular theory. In contrast, ethnography in system engineering predominantly takes the form of 'ethnomethodology' (Hemmings and Crabtree, 2002) and offers recognizable accounts of work situations and social interaction without such direct interpretation and categorization of user behaviour. The pragmatic use of ethnographic observational methods have been successfully shown in system development (see Hughes et al., 1994, 1995; Forsythe, 1999).

Contextual Design

Contextual design (Beyer and Holtzblatt, 1998) is one of the principal methods cited in studies that apply ethnographic observation in interaction design. It is an often-used technique in software engineering for handling collection, representing and interpretation of data of fieldwork. The method has seven parts: contextual inquiry, work modelling, consolidation, work redesign, user environment design, mock-up and test and putting it into practice. Information is structured in work models. The first part of the process, contextual inquiry, is an approach to the carrying out of an ethnographic study. It is best used in the early stages of development.

The most typical format of contextual inquiry is a contextual interview; a combination of observation, discussion and reconstruction of past events. The interview takes places during the work and tasks execution where possible. It does not follow a list of questions, thus contrasting with semi-structured interviewing. The inquiry involves observing large samples of work, asking questions and collecting artefacts. That information is then used to develop workflows and work descriptions.

Contextual inquiry is build upon four principles: *context, partnership, interpretation* and *focus*. The basic requirement of contextual inquiry is that it takes place in the actual *context* of the user/customer and that the work is observed as it is carried out. The goal is to recognize structure in tasks. *Partnership* involves making the inquirer and the user collaborators in understanding the work of the user. In case an observation does not fit observer's earlier understanding or a new structure seems to be identified, the observer interrupts and hopes to verify this with the user. Beyer and Holtzblatt aim for a mutual relationship of users and observer of shared inquiry and discovery of the user's work, with a close working relationship. *Interpretation* refers to the assignment of meaning to observations. Hypothesises about the observed events need to be generated and checked with users. *Focus* defines the point of view an observer takes while studying work, conversation needs to be kept on track to answer the questions a researcher set out to answer.

Examples of application of contextual design and contextual inquiry can be found in Väänänen and Ruuska (1998), Pasman (2003), Fouskas et al., (2002), Blechner et al., (2003) and Ko (2003).

Challenges in ethnographic studies

Authors such as Rouncefield et al., (1997) and Hughes et al., (1995) offer very useful advice in conducting ethnographic analysis and provide issues in quite a number of issues that come into play in observation. Some challenges in contextual inquiry are also reported by e.g. Väänänen and Ruuska (1998). Many challenges do however seem to pertain to observation in general; the right set of users has to be found, they have to be willing to be observed and a certain trust with the observed has to be established. Being observed can affect subjects' behaviour and some activities might be completely impossible to observe (Robson, 2002). Observation itself requires training; taking impartial notes, keeping focus and interpretation can be difficult. Setting the observation focus and keeping to that focus while being alert to observations that might be relevant to that focus may prove challenging when the researcher is confronted with complex environments and tasks. While it is important to notice the use of other agents or devices, at the same time the researcher has to deal with uncertainty due to non-observed actions of these other agents. Observational work (and potential such as video, audio data and gathered artefacts) can result in large amounts of more impartial data, while interpretation afterwards requires a large effort.

Forsythe (1999) is quite concerned about non-anthropologists dealing with such and other challenges while doing small-scale semi-ethnographic work. He names the limited generalization possibilities of small studies and stresses the importance of knowledge on ethnographic methods. Main challenges that need to be overcome in ethnographic studies are the pressures of time and the problem of scale. Ethnographic observation still takes a lot of time and effort, while only a limited of number of subjects in confined settings can be observed. This poses threats to depth and representativeness, which need to be overcome by e.g. maintaining a clear focus during a study (Hughes et al., 1995). Seeing ethnography as just another data gathering technique is warned against (see Löwgren, 1995) and Hemmings and Crabtree (2002) imply that 'interpretation should move beyond the gloss'.

Structuring and communicating contextual study results

One of the additional challenges of applying studies of context of use in system development is structuring and communicating the results of such studies in such a way that they can be easily used in the design process, a concern voiced for ethnographic studies in particular by for example Normark (2002). A number of frameworks for software development including requirements engineering for software systems are available that offer ways to structure data, for example contextual design work models (Beyer and Holtzblatt, 1998), scenarios and use cases can be used. Whether these or other tools suffice to structure gathered data, has to be assessed for the particular project at hand. Some information gathered in a study might not fit within e.g. workmodels or scenarios and can possibly not always be predicted beforehand. Such choices made in this study are elaborated upon in the methodology section of this thesis.

While such tools as the e.g. Contextual Design work models, scenarios and use cases may help to structure findings, they do not yet give an overview of the requirements for a system. They provide an overview of the current practices in the proposed context of a system, not a complete description of what utility is needed under what usability constraints.

Contextual requirements engineering for VR systems

Ethnographic observational methods are not completely unknown to Virtual Reality application development (e.g. Crabtree et al., 2002). However, endeavours to find out ore about the contextual requirements for VR systems are not often described in literature. In this thesis the use of such methods to study what challenges can be encountered in implementing a VR system for use in a day-to-day setting at the user's workplace is examined.

From literature background to VR development for the outside world

This literature background chapter shows the wide variety of choices that have to be made during the development of a VR system. The term 'VR system' refers to a wide array of systems with varying terminology. In development of such a VR system it has to be decided what type of VR system is suitable and what role presence and immersion should play. What evaluation method is suitable to evaluate a system's usability needs to be decided as well. Additionally problems have been identified concerning psychological factors, immersion and stereoscopy, navigation and orientation, physical effects and ergonomics and choices in (multimodal) interaction. Available guidance for developers on all of these issues is limited.

This chapter has by describing these issues provided a first insight in one of the research questions of this thesis; which usability issues make developing usable VR systems challenging? However, the identified issues from literature focus mainly on context-independent issues and it is likely that more usability problems can arise in specific contexts of use. This study aims to explore what which utility and usability issues exist for utilization of VR systems in a real world context. The use of contextual requirements engineering methods, described in the last part of this chapter, is explored here to identify context-dependent usability issues that might arise and to find out whether they can support development of VR systems for implementation in a non-experimental setting.

3. Case study

Little literature is available on the fit of VR systems within a prospect context of use. It has yet to be determined whether VR systems provide sufficient added value to warrant their use in spite of their potential drawbacks. Nor is much information available on the problems and choices encountered when developing a VR system intended for such a specific context. If VR systems under development actually meet user needs and constraints and what issues might challenge developers in achieving such a fit is unclear.

A case study has be conducted to see how the issues, outlined in the theoretical background section, come into play in developing with a VR system for a specific context of use. The issues and methods have been applied to a system that aims to support medical planning and diagnosis of abdominal vascular disorders; the Virtual Radiology Explorer (VRE). To explore issues that are encountered while developing a VR system with the intention to have it used in a day-to-day work context.

This section gives a brief introduction to the use of Virtual Reality in medicine and identifies additional issues from literature specific to development for medical VR applications. Literature has also been studied that might offer useful information diagnosis and treatment planning for vascular disorders and on requirements and constraints for systems that are aimed to support these processes. The VRE system is then introduced and subsequently this section poses the research questions of the conducted case study.

3.1. Virtual Reality applications in medicine

Medical systems are one of the main application areas of VR. Research projects into VR and medicine have been numerous and a wide variety of applications have been developed for both mental and physical health care.

In mental healthcare, applications have been developed for e.g. distraction for pain during burn treatment (Hoffman et al., 2001), aiding exposure therapy in the treatment of phobias or post-traumatic stress disorders by providing a simulated ‘threatening’ environment to the user (e.g. Klinger et al., 2002; Difede et al., 2002; Emmelkamp et al., 2001; Hodges et al., 1995).

For physical health care various steps in the medical process have been addressed; training of physicians, diagnosis, treatment planning and surgery itself. Teaching and training (e.g. Dinsmore, 1997), visualization, augmented reality, computer-aided surgery and surgical planning (e.g. Montgomery , 2001, Brown et al, 2001) are the main areas of medical application of VR. Augmented reality (e.g. Wildermuth, 2001, Banjura et al., 1992), telepresence and teleoperations (e.g. Kassell and Graves, 1997) have also been up and coming areas of development. Various clinical areas such as radiology, reconstructive surgery, dentistry, vascular surgery, and a wide variety of other specialisms have been involved. Literature offers a wealth of information on endeavours into developing VR applications aiming to assist medical diagnosis and surgical planning. Montgomery et al., (2001) provide an useful overview of research labs that develop medical VR systems. For older, but comprehensive lists of VR studies in the medical field see e.g. Waterworth (1999), Emerson and Weghorst (1994) and Satava and Jones (1999). Specialized conferences such as the "Medicine Meets Virtual Reality" (e.g. MMVR2003 highlights, 2003) show current developments and their existence is a sign of the great interest in medical VR.

VR systems aiming to support medical planning and diagnosis have been developed for various medical specialist areas, some applications of VR to diagnosis and surgical planning applications are outlined below.

Visualization
Visualization in 3D of medical data is the main functionality offered in many of medical VR systems. The primary source of data used is medical imaging, using data from e.g. MR and CT scans (El-Khalili, 1999). VR systems can assist diagnosis by offering the possibility for exploration of these 3D reconstructions. Abnormalities can be detected and explored in new ways. Zooming, viewing structures from multiple orientations, walkthroughs of e.g. the colon or arteries offer new possibilities compared with traditional viewing of printed 2D slices.
During interventions new possibilities for imaging of anatomy are envisioned, for example during keyhole surgery; where the anatomy of interest is not directly visible. Augmented reality can be used to show the patient's inner anatomy in real-time.

Surgical training
Using VR systems in surgical training offers a number of benefits (overview in Wildermuth et al., 2001). Surgical training using VR can be broader than regular training. All sorts of training scenarios with various sorts of patients, anatomical variations and operation environments that might not be encountered during training in the field, can be included in a system. Surgical skills can potentially be acquired faster by using simulations and performance of trainees can be objectified by automatic assessment.

Surgical planning
The promises of computer (and VR) assisted surgical planning include providing better surgical results, decreased time in the operating room, lower resulting risk to the patient and lower costs (Montgomery et al., 2001). The simulation capabilities of VR can be used to assist planning of medical interventions, e.g. to review results of simulations beforehand of proposed treatments.

3.2. Issues for medical visualization and planning systems

Various studies into novel ways of visualizing medical data for diagnosis and treatment planning are described in literature. Focus is mainly on using digital data obtained with medical scanners. Transitions from paper prints to desktop computer screens and e.g. VR systems all bring up issues that need to be addressed.

Van der Heyden et al.,(1999) give an overview of issues relating to the transition from using prints in viewing of scans (in their case MRI scans) to viewing of scans on a computer screen. Van der Heyden et al., list multiple reasons for this transition: hospitals or departments in remote locations want to exchange patient data, the potential of computerized medical image display systems and health issues relating to prolonged exposure to radiology prints. The difficulty with switching to computer screens is that the capability to view a very large image set simultaneously on one large lightbox traditionally used, is difficult to match on a small desktop computer screen. While the study of Van der Heyden is occupied with 2D visualization, the points made do illustrate issues surrounding the currently still prevalent way of assessing scan prints on light boxes and show what's important in viewing of scan images. These issues can also be relevant in developing 3D visualizations systems, especially when combining 2D and 3D visualizations.
Three main requirements are identified:

- Control; users need to be able to control over location, size, visibility and flexibility in creating 'groups' of images and control over these groups.
- Navigation; the user needs to be able to locate and relocate (groups of) images
- Detail-in-context; while screen size is limited, still both details (zooming) and the anatomical context of these details need to be available simultaneously.

Other requirements can be thought of and functionality such as measuring distances on scans, adding information to the scans, or changing contrast are not addressed in the study.

Additional design directions are given, but tested with a rather limited group of three radiologists and further research is deemed necessary.

Developing medical 3D scan visualization and VR systems for other medical uses such as surgical planning or teaching brings up more issues then when developing scan image viewer for a desktop computer. E.g. planning may require more information than can be gained from just looking at medical images, and developing VR systems or a surgical simulator brings up VR-related usability issues. El Khalili (1999) highlights a number of concerns in developing medical VR systems:

- 3D representation and manipulation of structures help in gaining better understanding of anatomy
- Although 3D representation is advantageous, physicians are not used to 3D interaction; traditional and 3D methods should be combined
- Generic data can be used in training, but patient specific data provides a variation that is useful in training
- Providing the physiology of the body is important for some medical simulators, but requires expensive modelling
- Accuracy of the anatomical, physical and physiological model of the organ is essential; however, interaction is the trade-off
- The interface between the physician and simulator should not be a burden
- Force feedback is important in some surgical procedures, such as minimally invasive surgery
- Evaluation of performance is important in planning and training.
- Providing distributed education and training applications adds to the value of the application, since it allows for group learning (this might be applied to diagnosis and planning as well to support possibly collaboration between physicians in clinical practice)

El-Khalili focuses mainly on training systems and also sums up requirements found in literature especially for surgical simulators:

- Model the anatomy of the organs; it is essential that the model is visually accurate and realistic
- Provide means of interaction with the objects in the environment
- The number of sensor channels, input/output sensory stimuli, provided exponentially increases the sense of realism
- Model physical and physiological properties of the organs
- Provide evaluation of user performance

It is important to note the difference between requirements for training systems (e.g. surgical simulators) and diagnosis and planning systems. Some of the issues outlined above might also pertain to systems aiming to support surgical diagnosis planning and not only to training systems, such as interaction and the modelling of organ properties and evaluation of results of a planned and simulated surgery are of interest in both situations. However, especially the striving to an increased sense of realism, might be useful in training, but less so in planning.

Amongst providing a number of interesting observations Montgomery et al., (2001) also illustrate that difference between training and planning. They present an interesting account of their experiences during a decade of research in the area of VR assisted surgical planning. They describe a number of systems and experimental use of these systems with real patient cases. The applications seem to focus mainly on cranio-facial reconstructive surgery. Montgomery et al., give a description of the systems and some of the lessons they learned during their research. Their insights include:

- Surgical planning is an abstract process unlike surgical simulation. In planning it is not necessary, nor desirable to conduct a complete surgery in Virtual Reality. Defining what is to be done and planning these actions is very different from actually performing these actions in Virtual Reality. A useful feature of Virtual Reality is the opportunity to overcome constraints of the real world and if using a virtual environment takes the same steps/difficulty/time as doing the real task, then the system is of limited use.

- Collaboration is important. Surgeons need to discuss patient data. Systems for planning should offer them to work together with patient data.
- Stereo visualization is useful for understanding complex anatomical structures (it is unclear if this has been concluded from experiments or from experience with systems)
- Offering high-resolution preoperative images from standard viewpoints used in medical imaging is useful, but offering interaction with this data is very useful to planning.
- Dynamic models of physiology and properties of tissue are useful in planning. Interaction with the data and seeing how tissue behaves would be desirable.
- Surgical plans sometimes need to be adjusted during surgery itself; availability of planning software during an intervention would be useful.

Montgomery et al., are confident that computer-based simulation, visualization and interaction are and will be very useful in surgery planning. They do however provide some notes on the adoption of medical planning systems. They find that surgeons are interested in systems that can help them improve patient treatment, but they are also critical of system shortcomings. The considerable cost of systems is an inhibitive factor; usefulness of systems and cost-effectiveness have to be validated; system benefits and ease of use need to outweigh its costs. It should be noted that every change from earlier work practices requires proof the new way of working is in someway more effective and/or efficient. Hinckley et al., (1994) stress that interfaces in medical planning systems should be easy to use for surgeons (in their study neurosurgeons). Surgeons do not have the time to spend a lot of time on learning to use a system. The economic incentive ‘deliver improved quality of patient care at a lower cost’ coupled with the high cost and time pressures of surgeons themselves will however make it more likely that money will spend on technology that actually makes planning more effective and efficient.

VR is deemed to be capable of delivering in accordance with many of the outlined concerns (El-Khalili, 1999). However, VR has its limitations and in many respects still mainly a technology in the lab. Also, it needs to be kept in mind that the conclusions of various cited authors might be specific to the specialist area they have been working in; various medical fields might differ in their constraints and requirements. Developing medical diagnosis and planning (VR) systems is complex and requires knowledge on technological, medical and human-computer interaction issues. It also has to be asked if VR is the optimal solution for a given situation. To obtain knowledge relevant to development of medical diagnosis and planning VR systems it is useful to gain insight on the actual context of use they will be deployed in; the setting might be very specialized and complex. This study explores the issues that are important to prospect users in such a specialized context of use. Focus here is on possible clinical use of a specific medical Virtual Reality system; the VRE, a system aiming to support diagnosis and surgical planning for vascular disorders. Following section introduces the VRE system, its functionality and development context.

3.3. The VRE

The Virtual Radiology Explorer (VRE) is a system for 3D visualization of medical scans, simulation and visualization of possible treatments for pre-surgical planning and training in Virtual Reality. The working prototype of the VRE system provides visualization of the results of blood flow simulation before and after a simulated bypass procedure. Its goal is to assist in the pre-surgical planning of abdominal vascular bypasses or to provide training. The potential users that have been considering during system development are surgeons, radiologists, medical lecturers and students. The VRE is now in a very early prototype state and has been developed as a technological testbed, therefore the system is not yet tuned to usability, contextual and user needs.

In the prototype virtual objects, including the vascular structures, are visualized in a virtual environment, which can be freely navigated by the users. Objects can be moved. The 3D reconstructions of scan data can be explored interactively and measurements can be conducted if necessary. GEOPROVE (fig. 10) Belleman et al., 1999) software is used for conducting measurements on objects in the virtual environment. ‘Geometric probes’ consisting of one or more (x,y,z) markers can be used to determine e.g. positions, lengths, angles, surfaces.

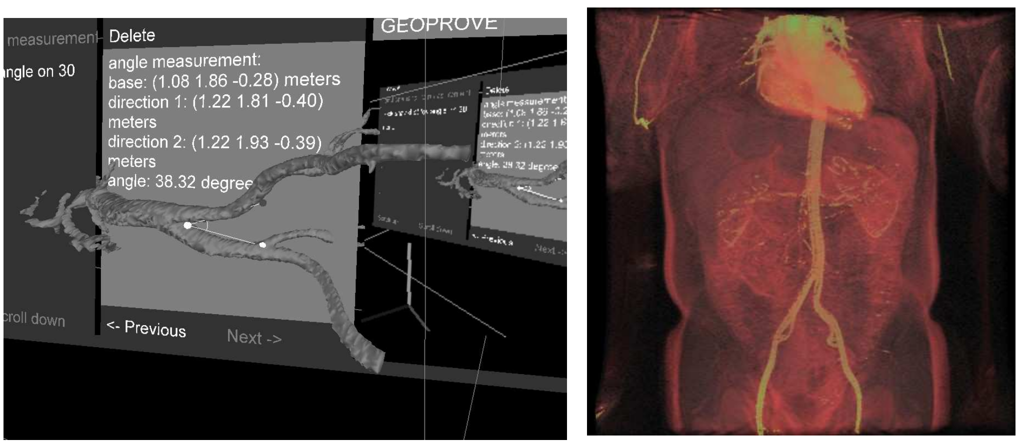


Fig. 10: Left: Visualization of a vessel within the first prototype of VRE, shown together with the floating menus of GEOPROVE, the measurement system applied within VRE (Belleman et al.,1999). Right: Exploration of patient data in VR (Zudilova and Sloot, 2003, image courtesy by D. Shamonin).

The simulation component of the system simulates the parameters of the vascular blood flow: velocity, pressure and shear-stress. Both the artery geometry and the simulated blood flow can be visualized in 3D as well as time (4D, see Belleman and Sloot, 2001). Simulation of flow (fig. 11) is visualized by streamlines showing flow direction, speed, pressure, in 3D and time. Colour denotes the speed of the flow (red for high velocity, blue for low velocity). The streamlines originate in an interactively movable plane consisting of particles. The number, shape and size of particles can be adjusted by the user.

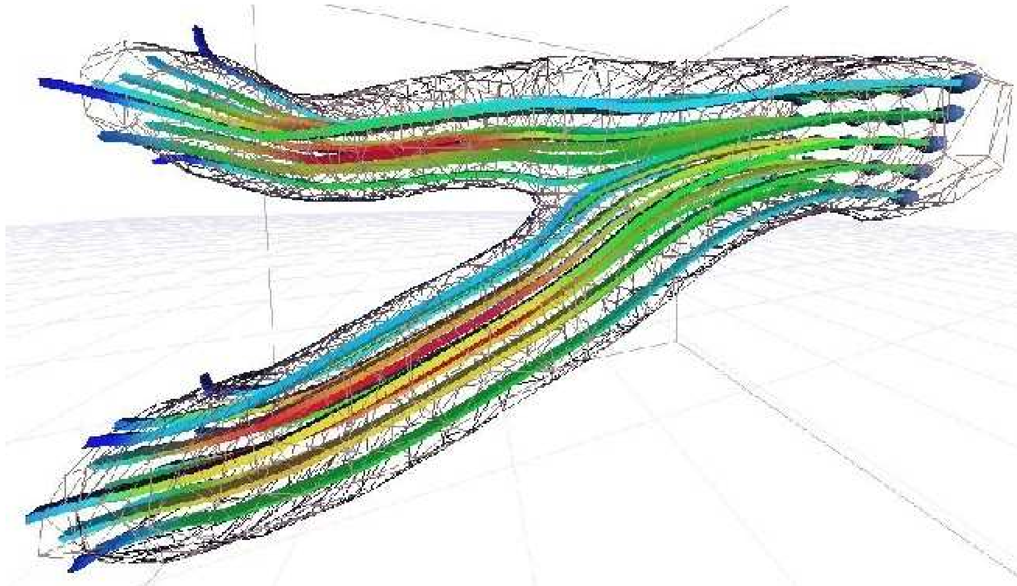


Fig. 11: 3D flow visualization with streamlines and particle plane on the right. (Belleman, 2003, image created by Hans Ragas).

Bypasses can be virtually tried out by putting a spline object on the visualized vascular structure or by drawing a connection. The resulting new vascular structure is then used as input for simulation of blood flow, enabling the user to see the simulated result of placement of the proposed bypass (fig. 12).

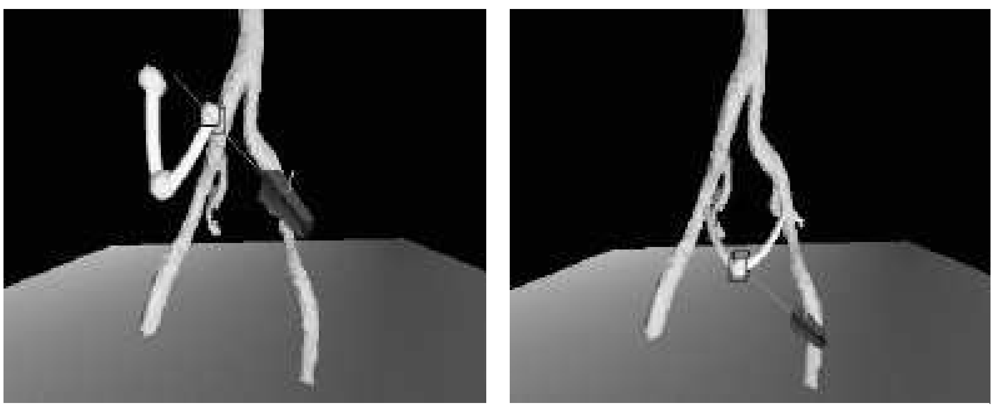


Fig. 12: Positioning of a spline object representing a bypass. Placement is carried out by positioning the spheres, the control points on the splines (Belleman, 2003)..

The VRE is an immersive application, which run now on semi-immersive DRIVE-system. The VRE can run in a fully immersive CAVE environment; a room-like life-size VR facility. Versions of the VRE for desktop computers and the personal space station (Mulder et al., 2003), a 'fish tank' VR system, are in development as well. The evaluated VRE version is tuned to the semi-immersive Distributed Real-time Interactive Virtual Environment (DRIVE) system (fig. 13, left). Pictures are projected on one large screen. The user wears shutter glasses (fig. 13, below right) to achieve stereoscopic vision. Separate images for the right and left eye are offered to create depth awareness. Synchronized to the screen images and at such high frequency human vision cannot detect, the shutters alternately black out vision for one eye, providing the other eye with a view of the concurrent image on the screen.

The VRE's interface is multi-modal; the user can interaction with the system in various ways. A magnetic tracker is used to track the user's head position, the view of the virtual environment is adjusted accordingly. Navigation and selection are achieved by manipulating a 'wand' space-mouse: three buttons, one joystick with two degrees of freedom and a tracking sensor, mounted in a single device (fig. 13, top right). Speech recognition can be used for spoken commands for interaction. For this, the user wears a wireless microphone. The speech recognition software is in an early prototype stage. Audio feedback is available for the spoken commands.

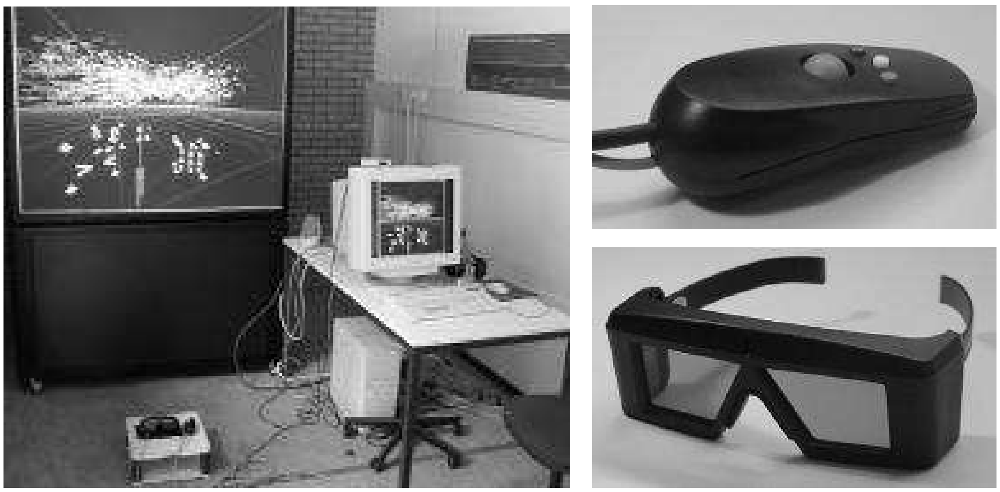


Fig. 13: DRIVE hardware, large screen, accompanying personal computer, wand, and shutter glasses. (images from UvA-DRIVE homepage, Belleman and Shukilov, 2003).

Grid developments are being applied to the VRE. Grid computation entails sharing of geographically distributed computational power. Connectivity between distant locations, interoperability between different kinds of systems and resources, and high levels of computational performance are some of the most promising characteristics of the Grid (Sloot et al, 2003). In the VRE Grid capabilities are used to for example offer connectivity between hospitals and apply distributed computational capabilities to manipulation of imaging data.

Extensive information on the VRE and the DRIVE can be found in Sloot et al. (2003), Belleman (2003), Belleman and Sloot (2001), Zudilova and Sloot (2003) and Zudilova et al. (2002), Belleman et al (1999), Belleman et al (2001).

3.4. VRE case study research questions

In section 1.1 of this thesis a general problem statement has been given in order to investigate usability issues for VR systems. The VRE has been used as a case study to explore these issues. This section will discuss the research questions that have adapted to this case study and extended from the general problem statement.

Even though the VRE has been developed in response to needs expressed by medical practitioners, the system has not been adapted to their needs, nor have these needs been studied adequately. The VRE has been developed mainly to test new technologies and simulation algorithms, as is the case with many VR systems. Its development has focused more on technological feats than on the user experience. However, when developing a system for actual end-users becomes a priority, an analysis of requirements, context of use and usability criteria for a system is crucial. Successful implementation of the application of a system for 3D medical visualization and surgical planning requires investigation of potential users, their tasks and their environments leading up to a complete interaction design. The VRE development process has not been ideal from a human factors perspective, since the goal for the system at first was development of technology and algorithms, and focus has now shifted to possible real use of their application.

The VRE prototype shows a number of usability problems, resulting from both the technical prototype stage of the system, as well as insufficient attention to the human element. An overview of potential areas where usability problems could arise is needed in order to inform further development of the system. An appropriate method has to be chosen to gain the most useful results of this usability evaluation.

Not only the system's usability in the laboratory, but especially usability issues for the system in an actual context-of-use and the fit of systems such as the VRE are of interest. The contextual requirements and constraints for the system have to be analysed in the real-life context of potential clinician user groups. Insight is needed into what tasks could be supported and what functionality is desired. It is yet unclear if a virtual environment would be applicable at all and if so, what type of environment and sense of immersiveness would be most supportive.

The focus in this study is on the possible clinical application of systems such as the VRE. Although some issues of VR in medical training are explored and possibilities for training can be identified, these are not extensively explored here. Focus in this case study is on context of use for medical diagnosis and planning.

This case study of the VRE aims to provide insight on the research questions (fig. 14) as presented in the problem statement of this thesis in section 1.1. For this case study a number of questions, which are specific to application of a system such as the VRE in a clinical context, had to be devised. These are listed below:

> What utility and usability issues exist for utilization of VR systems in a real world context?
> - Which issues make developing usable VR systems challenging?
> - Do VR systems provide sufficient added value to warrant their use in spite of potential drawbacks?

Fig. 14: General research questions of this study as discussed in section 1.1.

- *Does the functionality of the VRE adequately fulfil a user need or wish?*
    - *What user needs exist for:*
        - *stereoscopic 3D visualization of vascular structures?*
        - *simulation of blood flow?*
        - *simulation of the results of bypass placement?*

- *What are the contextual utility and usability requirements and constraints for visualization and simulation systems aiming to support diagnosis and treatment planning for vascular disorders in a clinical context?*
    - *How are diagnosis and treatment planning for vascular disorders currently conducted?*
    - *What are the relevant characteristics of potential users, their goals and tasks and environment in a clinical context of use, which should be taken into account in system development and implementation?*
    - *Is use of a system such as the VRE feasible in a clinical context of use; does it fit the prospect users and their goals and tasks, as well as the technical, social, organizational and physical environment of users that could potentially affect system use?*
    - *Is a choice for a Virtual Reality system appropriate for processes in this context?*

- *Which (unresolved) usability issues challenge developing the VRE as a usable VR system?*
    - *What are the usability problems of the VRE prototype independent of a proposed context of use?*
    - *What usability problems could potentially arise when the VRE would be used in a clinical context?*
    - *Are solutions available for these usability problems?*

The findings of this case study are aimed not only to be useful in the VRE's development, but also to be applicable and useful in a wider context. The study aims to add to the general human-computer interaction, Virtual Reality and systems development research communities by giving an overview of issues surrounding Virtual Reality usability, Virtual Reality usability evaluation methods and the contextual issues that come into play in developing a system to be used in medical diagnosis and planning. The study will try and identify obstacles that can be encountered while trying to develop a usable VR system for a specific context of use, including VR usability issues that warrant further research.

The study is also expected to offer insight into the effectiveness of contextual analysis and the applicability of ethnographic methods into investigating system requirements to support development of VR applications in general.

4. Case study methodology

A combination of methods has been applied in the VRE case study to gain insight in the issues posed. This section elaborates on the chosen methodology.

To gain more insight into the specific usability issues and requirements for the VRE, various methods are combined. To determine what the usability problems of the VRE prototype are independent of a clinical context, a heuristic walkthrough evaluation of the system has been conducted. To gain more insight into the context of use of the VRE and the subsequent requirements, constraints, need and feasibility for such systems an exploratory small-scale study has been carried out at a number of Dutch hospitals.

In this section the usability evaluation is addressed first, after which the context analysis is discussed.

4.1. Usability evaluation

For a first insight into potential usability problems of the VRE, heuristic evaluation has been coupled with a walkthrough of the system using representative tasks. In this case, it has been decided that an expert evaluation would be more adequate than testing the current prototype with users. Heuristic evaluation by experts is a reasonable quick and cheap method for evaluating usability of early prototypes (Nielsen, 1993). Staff working with the early prototype already identified a number of usability problems and were unsure whether it would be robust enough for an evaluation with end-users. Additionally, the cost and efforts of recruiting representative users of this very early prototype would have been considerable. The VRE cannot be easily transported and participants would have to come to the laboratory. This while the studied prospect users of the VRE are radiologists and surgeons who have little time to spare; getting them to visit and spend considerable time at the VRE's development laboratory would be difficult. In its current state of development the VRE; in which the system has not yet been adapted to their tasks and is constantly changing to considerable degree, the readiness of participants to come in multiple times and evaluate the system in this state would be doubtful. The goodwill of the participating medical professionals cannot be exploited for a very small gain in system development, without concern for the value they have in their daily workplaces. It is important to retain the willingness of available contacts to participate in future developments of the VRE and not to expect goodwill to be unlimited.

Additionally, user testing is most applicable with ample knowledge about the prospect context of use of a system. Dix et al.,(1998) for example point out that usability metrics rely on measurements of very specific user actions in very specific situations. It should be understood why and how particular usability metrics are applicable in certain situations and how they provide for greater usability for real users. Before a complete usability evaluation with users can be a meaningful instrument, it has to be known who these prospective users of a system will be, what tasks a system should support, and what the context(s) of use will be. This is not sufficiently the case in the early design stage as the VRE is in. Focus of development has been on technological development and a complete analysis of the context of use has not been conducted. Therefore this evaluation has focused on finding the main usability issues of the interface of and functionality of the VRE prototype. The observations relating to the prototype stage of the VRE, combined with the relative low cost of carrying out a heuristic approach makes qualitative heuristic evaluation of the system a more viable option at this stage of development.

This first evaluation does not yet give an overview of issues that might come into play outside a lab setting and it is acknowledged that subsequent inquiries into that context of use and evaluations with users should be performed. The evaluation does however yield insight in the relatively context independent issues that come into play in using a VR interface for certain applications. Usability issues, identified during the evaluation, also give insight into issues related to the future context of use of the system that need to be studied for further development.

One individual predominantly carries out this evaluation; this is unlikely to result in finding all usability problems (Nielsen, 1993). In this early stage evaluation, this is deemed acceptable. It is not the aim to find every single small usability problem, but to identify major areas and themes where usability problems might arise and where additional information on the context of use of the VRE is needed. It is not claimed that small individual usability problems cannot have very serious consequences on overall systems usability and acceptability. However, in the case of the VRE, a complete redesign of the interface is not unthinkable and this approach might be more efficient in this case.

Specific heuristics for VR systems are used in the evaluation. Bowman et al., (2002) in their overview of usability methods applied to virtual environments, notice that a lack of well-formed guidelines and heuristics for user interface design and evaluation of virtual environments make heuristic evaluation challenging. Indeed a set of guidelines has to be devised in this study. Valuable insights and guidance into heuristics and specific evaluation of VR applications as been described in the theoretical background section of this thesis have been taken from a number of authors; Kalawsky (1999), Steed and Tromp (1998), Kaur (1999), Suttcliffe et al., (2000), Neale and Nichols (2001), Willans et al., (2001).

The system walkthrough is performed to also gain insight in possible issues that might not be addressed by the set of heuristics. Multiple walkthroughs are performed with considerable time between them, both to address possible oversights of earlier evaluations and to address changes in the VRE prototype that is under constant development. The evaluator walks through all functionality of the VRE. Specific tasks representative of the VRE functionality and the expected tasks to be performed with the system are also performed by the evaluator to test the system. Tasks include exploring a modelled artery structure and blood flow simulation results to diagnose the condition of the vessel and the placement of a virtual bypass on a virtual vascular structure and exploring its results on blood flow. The basic cognitive walkthrough questions as described by (Dix et al., 1998) are used:

- *Will the users be trying to produce whatever effect the action has?* Are the assumptions about what task the action is supporting correct given the user's experience and knowledge up to this point in the interaction?
- *Will users be able to notice that the correct action is available?* An interface item, like an option in a menu, needs to be visible to the user. Apart from if the item is understandable, the user should be able to find the item.
- *Once users find the correct action at the interface, will they know that it is the right one for the effect they are trying to produce?* An interface item not only needs to be visible, but it has to be understandable as well to the user.
- *After the action is taken, will users understand the feedback they get?*

The usability problems that are found are ranked according to severity following the approach of Steed and Tromp (1997) following Nielsen. The severity ratings are given according to the reasonable combination of three factors:

- Frequency of the problem; how common the problem is.
- Impact of the problem if it occurs; how difficult the problem is to overcome for users
- Persistence of the problem; can a user see the problem as a one-time problem which they can overcome if they know how to deal with it, or will they be repeatedly bothered by it.

It has to be kept in mind that not a complete overview of all usability problems of the VRE and their rating will be given in this thesis; only the main issues and problem areas will be addressed.

4.2. Contextual analysis methodology

Usability evaluation of a prototype in a laboratory setting does not yield insight in the actual context of use of a system. The research questions of this study focus on the requirements for a system such as the VRE in a clinical context. Focus herein is on the assessment of medical scan images and blood flow in the process of diagnosis and treatment planning; the main type of tasks the VRE has been developed to support. In order to ensure that VRE functionality will be developed supporting real-life demands, the choice has been made to conduct an exploratory study as a first step to investigate the daily working context of two of the end-user groups; radiologists and vascular surgeons. A contextual analysis is carried out combining exploratory interviewing with observational methods adapted from ethnographic research.

The ethnographic approach in this study combines semi-structured interviewing and observation. Vascular surgeons and interventional radiologists are interviewed to get an overview of the work processes surrounding the tasks of diagnosis and treatment planning for vascular disorders. Observation sessions are carried out to gain detailed understanding of the tasks and the context in which they are performed. An interview is planned to precede a first observation session with a participant. An opportunistic approach is taken in that every opportunity to observe real work carried out is exploited. Each visit to a hospital results in a narrative report of that particular visit alone and the information gathered.

The study can be mainly categorized as *quick-and-dirty ethnography (*Hughes et al., 1994) for a first overview of the work domain, but also takes after *evaluative ethnography* in checking implications of the VRE's design for fit within the proposed clinical context of use. This study is similar in its application of ethnographic methods as earlier studies in software development (see Forsythe, 1999) and Virtual Reality application development (e.g. Crabtree et al, 2002) borrowing elements of Contextual Inquiry techniques (Bayer and Holzblatt, 1998).

Apart from combining techniques while appreciating their strengths and weaknesses, this study follows other general guidelines for data gathering for requirements engineering that are offered by e.g. Preece et al (2002). All stakeholders groups are involved, and focus is on their needs. While the focus is on the two potential direct user groups, surgeons and radiologists, needs of other stakeholders (e.g. patients, other hospital staff) that might be identified are certainly considered as well. Multiple participants that are members of possible user groups are involved in the study. Preece et al., also recommend that data-gathering sessions should be supported with suitable props, such as task descriptions and prototypes if these are available. Specific props are not used during site visits, since the goal is to gain insight in the current ways of working, to explore possible wishes and needs of participants and not to directly use the VRE prototype herein. The VRE prototype is not shown to the participant, this is deliberate to gain a more system-independent insight in the needs of the participants in the processes of diagnosis and treatment planning. The system's functionality is however described to participants, e.g. if necessary to gain their participation in this study, on participants' request or to gather more adequate information on e.g. system feasibility. Preece's recommendation to run a pilot is followed, be it by adjusting interview questions and observations when these do not result in adequate data.

For further development of the VRE possibly other studies with other techniques such as user testing and more participatory design techniques will be used, this is however beyond the scope of this thesis. The following section will first elaborate on the methodology applied in the interviews, after which the specifics of observation sessions will be discussed.

4.2.1. Interviewing

First, a round of semi-structured interviews is carried out to get an overview of the work processes and issues involved in the tasks of the complete process of diagnosis and treatment planning for vascular disorders. This first round of interviews is in keeping with advice of for example Wood (1996, as quoted in Hackos and Redish, 1998) who suggests starting out contextual inquiries with semi-structured interviews with key informants before observing their (an others') work. Preliminary interviews are useful to understand the general context of the work, vocabulary, and general work issues before carrying out observations that might not be understood without background information. The main goals in carrying out semi-structured interviews are identifying (sub)tasks in diagnosis and treatment planning, identifying actors in this process and potential users and stakeholders of the VRE. This way, candidate tasks which the VRE might support can be identified and can give an indication of which tasks should be studied in the observation sessions of the study.

Interviews in this case are preferred over a survey type inquiry and semi-structure interviewing was chosen over structured interviews to gain this flexibility. Not enough information is available about the users, their tasks and environment to use a fixed list of questions. In this exploratory phase it is important to explore issues that come into play in diagnosis and treatment planning, indicated by participants themselves. However, a list of questions was used to focus the interviews. The interview questions are adapted when new points of interest arise from earlier interview results. Later interviews are also used as an iterative tool for verifying and comparing earlier study results. Extra questions from developers are integrated in later interview versions when necessary. Questions of the various versions of the interview can be found in Appendix 1. It needs to be kept in mind that due to the flexible nature of the interviews, additional questions have been asked. Due to acute medical emergencies and subsequent acute participant calls, some interviews may not always be finished completely. Such information is recorded in the individual interview reports, available with the author on request.

The interviews are carried out in a one-on-one setting, at the work location of the subject in a room of his/her choice. This minimizes discomfort for the subject and permits them to stay at their departments; which is important for medical staff in the case of acute patient emergencies. Interviews are taped after permission from the participant and added notes are made on paper. If dedicate interview sessions are inconvenient for participants, interview questions are whenever possible asked during observation sessions. Interview structure and questions are devised respecting guidelines as put forward by e.g. (Preece et al., 1998). Interviews start with an introduction, a warm-up with easy non-threatening questions, the main session, followed by a cool-off period with some easy questions and a closing section to thank the interviewee. Questions are clearly worded as short and easy as possible, non-leading and in keeping with the goals of the study (see appendix for interview questions).

Participants are asked about their responsibilities, activities and information used concerning the tasks of diagnosis and treatment planning. Information is gathered on e.g. aspects brought up by Hackos and Redish (1998) that can be highlighted in an interview trying to understand tasks an actor is carrying out. The process analysis aspects are; time, timing, triggers, the steps that constitute the task, who executes task, the information that is used and what information comes out of the task, when the next task is executed and by whom. Current bottlenecks and high-risk elements in the users' task set are assessed in order to gain understanding of where system support could be useful. Additional specific questions are asked to address issues surrounding the simulation and visualization capabilities offered by the VRE system. The current use of 3D data is evaluated as well as wishes participant might have for future medical visualization tools. The current use of and future intention towards use of blood flow simulation data is explored as well.

4.2.2. Observation

To gain greater insight in the working situation and tasks of the participants observation sessions are carried out. One of the goals of this study is to give an overview of the prospect context of use of the VRE. This includes all the aspects of context of use as defined for this study: real-life tasks, technical, social, organizational and physical environment of users that could potentially affect system use. Only using interviews is not enough to gain a complete understanding of the context of use of a prospect system. When only using interviews external validity would suffer from the same problem as other techniques that do not directly observe behaviour suffer from; behaviour and attitudes are not necessarily accurately reported. The main point of contextual analysis is seeing work being carried out in its context. It is recognized that tasks, use of systems and artefacts and interactions are difficult to analyse without actually having seen participants carry out their daily tasks, interacting with each other and using systems and artefacts.

The information gathered in the observations includes, but is not restricted to data concerning: current workflow, tasks and artefacts, prospect user characteristics, collaborations between prospect user groups, current scan visualizations, tools and software used and the social environments. It is made sure that observation includes all tasks in the process of diagnosis and treatment planning for vascular disorders by (vascular) surgeons and radiologists as identified in the first interviews. An opportunistic approach is taken by the researcher, where every chance of seeing work being carried out is used. Unplanned observations are unavoidable in a context the researcher does not yet know well and are welcomed.

The observations are carried out in a manner both useful to the researcher and the study's goals as well as appropriate to the sample environment. Discomfort to participants (and possibly other parties, such as patients) is minimized as much as possible. The observations take after Contextual Inquiry interviews, with the difference that Contextual Inquiry interviews are, as the name indicates, interviews and questions are concurrently asked during observation. In this study, while questions are asked, the emphasis lies more on observation than in a Contextual Inquiry situation. This has to do with the medical setting in which Contextual Inquiry may not be the most appropriate strategy. In this study, meetings or consultations with patients for example are observed in a fly-on-the-wall fashion to not disturb the participant and their patients. Questions are asked after the meeting to the individual subject. The agreement to observe has been made with the subject, provided that the time schedule permits. During individual tasks questions are asked more freely, but often after the fact to not disturb concentration during important tasks.

A typical data collection session starts with an introduction and an explanation of the research and its goals, an explanation why and what the procedure is for the participant's observation. The session program is discussed with the participant and adjusted to minimize the disturbance to the participant's work. Additional departments are visited when the opportunity is offered or created by asking for contacts and extra information. Observation is carried out by an individual researcher. Observations and analysis of observations are verified when possible with participants themselves and otherwise with domain experts. Notes are taken on paper. Physical lay-outs of rooms and lists of tools present in the rooms are noted. Photographs are taken after asking permission from the subject(s) when time and confidentiality permits. Information gathered in an individual session is written down in an electronic document.

Additional information pertaining specifically to the VRE system, such as possible requirements for 3D visualization and blood flow visualization and simulation, is gathered as well, both in interviews and observation sessions. An overview of the gathered information can be found in the results section of this thesis.

4.2.3. Structuring and presenting results

This study should result in structured reports of findings and recommendations for future development of the VRE. The findings will be structured in a number of ways, which are elaborated upon in this section.

For each interview and observation session an individual report listing findings is made. These individual interview and observation reports are analysed to provide a more generalized overview. Borrowing terminology from the Contextual Design method; a 'collated' overview of findings is provided. CommonKADS worksheets and Contextual Design work models were originally to be used to structure gathered contextual information. The worksheets and work models offer a long list of issues and characteristics of users, task and context that can be relevant in considering a context of use in requirements engineering.

Although the CommonKADS method (Schreiber et al, 1998) has been specifically developed to cater for development of knowledge based systems and does not specifically promote ethnographic studies, the method does offers useful guidance on studying contextual factors that might affect development and implementation of other types of systems as well. The CommonKADS models offer a structured way to present information gathered in the requirements engineering process that can offer insight in user characteristics, their needs, tasks and environment; information that can be gathered in a contextual study such as this one. However, the work sheets are rather specific to information exchange and knowledge systems and some of the characteristics made them less relevant to the purposes of this thesis. In contrast, Contextual Design work models are often used for structuring ethnographic data. An example of a disadvantage in communicating with the Contextual Design models is the use of a non-standard notation, this could be countered by using a standardized modelling language, such as UML, where possible and trying to design the models as clear as possible.

Not all elements of the two methodologies can be applied in development of a system such as the VRE. Elements of the methodology have to be adapted and it has to be taken in consideration that specific contextual influences that are relevant to specific kinds of systems, in this case VR systems, are not addressed by the CommonKADS and Contextual Design methodology. These models have been partially completed, however, in hindsight they provided little additional value to the listing of findings by theme and some aspects like e.g. providing maps of work space went beyond needs for the contextual study. Including the models and work sheets in this thesis goes beyond its purposes.

Scenarios and use cases are used as examples trying to explicate more abstract findings and for giving developers a flavour of the daily work and context of particular users of a proposed system. They are not used as the basis of requirements engineering, but used as a communication tool. Additional information about requirements for system functionality and specific issues such as requirements for blood flow simulation is recorded separately. This information is grouped in themes and reported in a narrative fashion or in diagrams.

To communicate results informal conversations are carried out between the researcher and the VRE development team. In this way, insight can also be gained on what questions the team might have and what information it might additionally need. Presentations along the way are useful to inform larger groups of people of preliminary results and to gain their input on the ongoing study. Sub-reports are used to gain such input during the study as well, both for the development team and the researcher.

5. Results

This section discusses the findings of both the VR usability evaluation in the laboratory and the contextual study. First the results of the heuristics and walkthrough usability evaluation of the VRE prototype itself will be discussed. The results of the contextual study's interviews and observation session will be presented afterwards, including discussion of the applicability of a system such as the VRE in a clinical context.

5.1. Results of the heuristic evaluation of the VRE

The walkthroughs and heuristic evaluation of the VRE has helped to identify a number of usability problems concerning the existing VRE prototype. This section discusses these problems. An example is offered of an interaction episode with the VRE. This example is meant to illustrate some of the encountered usability problems and give the reader a sense of the current interaction with the VRE's interface. This section will further elaborate on the main usability problems encountered.

It should be noted that findings from this heuristic evaluation are subjective due to the limited number of evaluators; the evaluation is predominantly carried out by one researcher individually. The evaluation can merely be seen as insight into the potential usability problems that may arise. However, an in-depth qualitative analysis of potential problems that might arise is useful to guide further research and development. Identifying potential user problem areas in the interface design can highlight the need to address these issues in development of the VRE and other VR systems. It can also provide insight in possible principal issues that have not yet been satisfactory addressed in VR research. To eliminate the subjective aspect of this study, further research will include quantitative analysis of the issues highlighted during the heuristic evaluation and user profiling (Zudilova and Sloot, 2003).

It is not claimed, nor has it been the goal, that all individual usability problems of the current VRE prototype have been identified, nor that all identified problems are listed in this document. It is beyond the purposes of this thesis to list all found usability problems of the VRE. A selection of usability problems of the VRE that are particularly relevant to the VRE's further development and future design choices are highlighted in this section.

Issues like for example the effectiveness of the way in which blood flow is shown with lines are not discussed here. Usability of menus and commands is also only limitedly discussed here. It goes beyond the purposes of this thesis to discuss problems such as colour schemes, or learnability and recognizability of menu options and commands. While acknowledging that these issues can be important to usability of a system, the purpose of this results section is to give the reader insight into the more high-level issues regarding usability of the VRE. Some of these additional problems are however mentioned in the example of an interaction episode with the VRE.

Difficulties in starting up the DRIVE system and starting the VRE application, are not included in this evaluation either, only the VRE application itself is discussed. It is acknowledged that these procedures however are important to overall usability of the system, and are currently far from ideal in regard to time and complexity.

Issues resulting from the prototype stage of the VRE and technological challenges of VR development in general are discussed first in this section. Some problems however also seem to result from using an semi-immersive virtual world type application for this system and resulting navigation challenges. These seemingly paradigm-related problems will be highlighted afterwards.

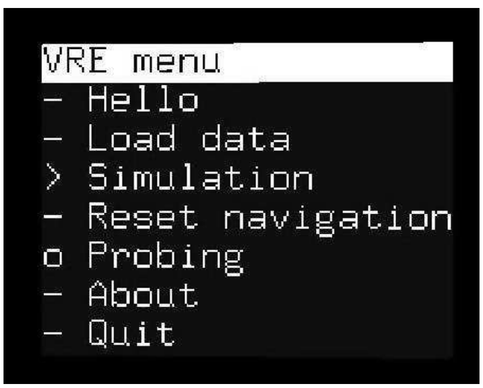
5.1.1. Interacting with the VRE

Interaction with the VRE takes place via either point-and-click menus via the ‘wand’ space mouse, via voice commands, and head-tracking. Items such as menus, commands and accompanying functionality of the evaluated prototype are explained in this section.

Interaction techniques within the VRE

Interaction with the VRE takes place via head-tracking for orientation and viewpoint changing within the virtual environment, for navigation, point-and-click interaction and voice commands for e.g. object selection and manipulation (e.g. changing an object’s colour) and system tasks (e.g. starting a simulation). Point-and-click interaction within the VRE is achieved by pointing with the space mouse ‘wand ’and pushing its buttons. Fig. 15 shows the main menu of the VRE version as available during the usability evaluation and explains the functionality associated with the menu options. The main menu and the voice commands would be incomplete for a system actually used for diagnosis and tram and serves mainly to facilitate testing of functionality and development of the VRE. For example, there is no way yet to load other data than the example artery structure and the standard example blood flow simulation, nor can interactions be saved. It is also noted that important features such as ‘undo’ are not yet included in the system.

VRE main menu



Menu options:

Hello: test function; selection results in the audio feedback with a voice saying “hello”
Load data: loads the test data and shows the accompanying vascular structure
Simulation: offers a submenu with simulation options; **start, pause, continue** the blood flow simulation and visualization. Only one standard example simulation is run in the prototype.
Reset navigation: resets the navigation to the standard position and orientation
Probing: opens the GEOPROVE measurement tool as a window within the VRE environment (for an extensive discussion, see Belleman et al., 1999)
About: gives version information
Quit: closes the VRE application

“>”: selection of this menu option opens a submenu
“*o*”: selection of this menu option opens a new ‘window’ within the application

Fig. 15: The VRE main menu in.

Example of interacting with the VRE

Appendix 2 shows one of the test tasks that where used in the heuristic and walkthrough evaluation. Three main tasks have been used for this usability evaluation:

- Inspecting an artery
- Placing a bypass between two veins of a vascular structure
- Running a blood flow simulation

The basic task of visually inspecting an artery is discussed in the example. There are alternative actions for some of the interactions and the example shows only one path of interaction through the VRE for this particular task. All interaction modalities are used in the example and a number of usability problems that have been encountered during interaction are listed. Not all ‘errors’ and problems might occur every time the user tries a particular action. For example, many problems have arisen due to faulty interpretation of voice commands, a correct -or another incorrect- interpretation a user e.g. pronounces a command differently.

Interaction methods and usability problems of the VRE

A number of problems in using the functionality of the VRE resulted from limitations of the current state of technology used for the interaction methods of the system. The different input modalities were in an early prototype stage and both the tracking system and voice recognition system posed the user with usability problems. Ergonomics of input devices is a concern as well. A number of these problems are discussed below.

Tracking

Problems were apparent in the magnetic tracking system used. The magnetic properties of an assortment of items in the testing lab, ranging from file cabinets to metal in the walls, resulted in an intense jitter which made pointing and selection in the VRE very difficult. Some lag in real-time feedback resulted from limitations of computational resources in for example running and visualizing extensive simulation of blood flow. The jitter and lag at times rendered working with the VRE very difficult. Selecting a menu option could be very challenging and e.g. placement of markers in the artery structure very imprecise. This section focuses on context-independent usability issues, but it needs to be noted here that an added concern for the choice of this magnetic tracking system exists for the proposed medical context. The reaction of the system to metal and magnetic parts in its environment might pose serious challenges in a medical environment containing a lot of equipment. Especially in e.g. radiology departments with magnetic scanning equipment this might be a serious concern. This, while precision in e.g. medical measurements might be critical.

Voice recognition

The voice recognition system did not always correctly interpret commands from all users and not all relevant possible synonyms were included. In a ‘real’ context of use speech recognition might be less desirable. The recognition ideally should be robust enough to handle environment noises and interruptions. Some users might find using speech recognition embarrassing (talking to a screen, being heard by others) or might be uncomfortable when others can hear that they make mistakes (irrespective whether these are actually a consequence of a system design problem) and the system ‘does not understand’. This would certainly impair the social acceptability of a system. However, voice recognition could be a good alternative for users whose hands are occupied or find manual interaction in VR cumbersome. Context of use and user characteristics might be very important factors in usability of specific modalities, as for example speech recognition.

Input modalities

Information on what kind of interaction modality is suitable for a certain type of interaction is yet scarce. The VRE offers a number of input modalities, not all individually suited to the same type of interaction. It is for example rather inconvenient to move an object in the VRE environment by using only voice interaction, repeating the ‘move left/right’ commands, while selecting an object with the space mouse pointer and pointing to the desired final location, combined with a ‘move here’ voice command works rather more conveniently.

What (combinations of) interaction techniques are appropriate for specific actions is yet unclear. Perhaps it can be argued that at this time, a system design should, whenever possible, provide the user with various interaction methods and support achieving various actions using any of these offered interaction methods when desired. While the VRE a this time somewhat supports this, some inconsistencies are present which should be corrected in a final system and research is necessary on determining what interaction techniques are suitable for particular situations.

Input via the wand, pointing a beam on screen to objects, poses questions as well. Even if a user would be immersed in the virtual world, the pointer used within the VRE to manipulate objects, seems a unnatural way to manipulate an object and seems to create a perceived ‘distance’ from the virtual world. Using a beam as a pointer, virtual objects cannot be ‘grabbed’, but have to be pointed at from a distance; direct manipulation is not accommodated. Alternative input devices mimicking the virtual objects could possibly offer a solution, in this case for example, small physical models of the abdominal aorta could be envisioned. However, in a virtual world comprised of many types of objects this might not be desirable and pointing to distant objects might still be necessary. Research on new possibilities for input devices and user preferences in this area is needed and experiments are being conducted with models of virtual objects for interaction (e.g. Mulder and Van Liere, 2002).

While the encountered problems above can be expected in an experimental set-up, their occurrence does show a immaturity of both technology used for these interaction techniques and guidelines for their use. Both will have to be addressed before systems implementing these technologies can be satisfactory used.

Ergonomics
A hardware-related issue is the physical discomfort users can experience while working with the prototype, not only in regard to simulator sickness, but also to the ergonomic qualities of the wand, headgear and the shutter glasses. Putting on the equipment requires quite some effort on the user’s part and wearing it can be rather uncomfortable. While more comfortable equipment might be available, ergonomic improvements are an ongoing quest in VR related research and equipment has not been found ideal. The physical positions users need to assume can at times make using the VRE a rather tiring physical experience. A user needs to be standing upright to use the VRE, and constantly hold the space mouse in front of him/her. Additionally, pointing and tilting the space mouse can require quite uncomfortable positions, not an uncommon problem for VR applications. It needs to be pointed out that while this is mainly a context independent problem, in a clinical context such ergonomic problems, especially simulator sickness, can be very serious. If the system were to be used for example by a surgeon to plan a procedure during a regular work day this could be unacceptable. A nauseous, dizzy or disoriented surgeon with an unsteady hand might be very detrimental to quality of his/her decisions and actions during e.g. a surgical procedure that might follow system use/

Even more important than these problems above, is the choice for a certain interaction paradigm. The offering of an explorable virtual world has consequences for the usability issues that are encountered. The following section will go into these issues and the questions whether the VR paradigm is suitable for an application such as the VRE.

5.1.2. The VRE and the VR paradigm

Some major usability problems seemed to result from the choice of a VR paradigm for the VRE application. These usability issues may be important in considering (not) choosing a VR type application and are discussed here.

Systems such as the VRE primarily aim to support inspection and manipulation of virtual objects (in this case vascular structures) rather than exploration of and presence in a virtual world. The free exploration possibilities offered by the VRE, while system functionality focuses on 3D objects and not world exploration result in a number of usability problems.

First of all navigation issues can be identified. Due to the exploration possibilities offered in the virtual environment, users could get lost in that environment and lose sight of the vascular structure object of interest. This while the environment around the structures does not yield any information for a diagnostic or planning task. Interesting also is the absence of any indication of where in the patient’s body the shown artery volume is located. The focus of the VRE is now on abdominal aorta (with a main vessel in the abdominal area, splitting in two vessels to the leg), the artery object is shown outside of a bodily context and also shown tilted as if a patient would be lying down, adding to the disorientation.

The combination of 2D and 3D objects in 3D space has posed some problems as well. Virtual menus have been modelled as 2D objects in a 3D world. This strict adherence to the paradigm of a virtual environment poses problems in using the system. Just like other virtual objects the menus could be moved out of sight, or could be explored from the side and back. This made selecting a menu option into an elaborate task that could include extensive navigation of the environment while trying to locate and adequately position and orientating the menu. Fig. 16 illustrates similar problems.

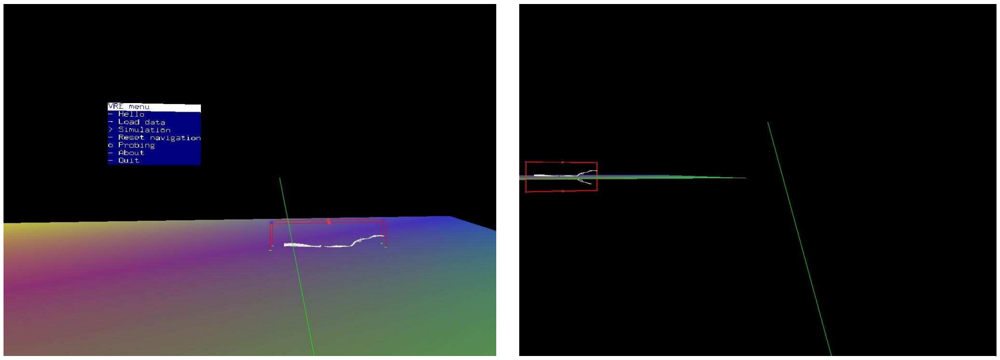


Fig. 16: Two examples of problems in exploring the VRE's virtual environment. On the left screenshot, most of the screen space actually shows the virtual environment to the user, which offers little information to the user. This, while the object of interest, the artery volume, has been moved through the plane that forms an orientation item in the environment and is barely visible. The plane now obscuring the actual object of interest. On the right screenshot, all items of interest are almost completely outside the user's view, offering the user a view of mainly 'virtual space' with objects of interest in the far distance, making them hard to see. Inspecting these objects will now require extensive navigation in the virtual world.

Research into more optimal ways of offering menus in virtual environments could be of value. This also counts for ways of offering other 2D information that might be of use, e.g. 2D images or text. This is especially relevant for medical applications, which aim to support clinicians who also use 2D scans.

However, for this system, the evaluation indicates that the problems that have resulted from choosing a complete VR system in the definition as used in this thesis ("a system offering computational simulation, 3D visualization and an artificial environment that the user can interact with in real-time") is not ideal. The unique properties of VR; the possibility of exploring a virtual world and the offering of the experience of presence in such a virtual environment do not seem to offer distinct advantages for this application. The offering of an environment, including making all objects 3D objects that could be freely moved, in this case even posed major usability problems.

Solving these problems of free navigation and combining 2D and 3D in virtual environments might indicate a necessity to switch to another paradigm, perhaps moving away from VR or moving to hybrid variations. For example, using fixed menus and restricting 'navigation' to 'exploring an object' within a desktop type application might solve some navigation issues. Certainly such hybrid forms of interface paradigms are possible and known, e.g. CAD design systems, or other desktop 3D applications; categorized by Marsh (1998) as fuzzy boundary systems.

In general, more information is needed on requirements, context, usability problems and the benefits of using VR for certain applications.

5.2. Results context analysis

After usability evaluation of the VRE interface, the need for more information on the context in which such a system might be used and required functionality has become even more apparent. To find out more about the utility and usability requirements for a system aiming to support diagnosis and treatment planning, a contextual study has been conducted. This section will elaborate on the results of this study. Characteristics will be discussed of potential users, their goals and tasks and environment in a clinical context of use, which should be taken into account in system development and implementation.

5.2.1. Participants

Medical staff from nine hospitals have participated in the study. The involved hospitals were geographically spread within the Netherlands. They varied in size and were both academic and non-academic. Fourteen specialist participants, of which seven (interventional) radiologists and seven vascular surgeons, have individually been interviewed and/or observed. All individually interviewed and observed radiologists and surgeons were male and over forty years old. Two female radiology technicians have been observed during post-processing of scans and two other technicians briefly have demonstrated and offered opinions on the software they use. Table 2 provides an overview of the hospitals and participants. Additional (varying from one to about fifteen per activity) radiologists, vascular surgeons, other specialists, technicians, nurses and assistants have been observed during vascular meetings, surgery and informal discussions on scan assessment. Some of them have answered questions, or have provided comments, but they have not been extensively individually interviewed and are not counted as individual participants.

The participants for the interviews and observation were selected using contacts in the medical profession of the research group and by using these contacts as starting points for recruitment. Taking such an opportunistic approach and using participants as starting points to recruit more participants is useful, but can result in biased sample, and participants may be self-selected because of their interest in technology (Forsythe, 1999). In this study this effect was countered by recruiting participants from different hospitals and using different starting points, but the effect cannot be ruled out completely.

The preliminary interviews have been carried out with three vascular surgeons and one radiologist before any observation was started. The questions in the interviews have been asked to the other participants when the information was not clear from observations. The questions have been asked either as a complete one-on-one interview or as questions during observation sessions at unobtrusive moments. The choice for one on one interviews or asking questions during an observation session has depended on the preference and time constraints of the participant.

N	Hospital size	Participants: Surgeons/ Radiologists/ Technicians (**I**nterview / **O**bservation**)	Activities observed	Regular scan assessment (prints/screen)
Non-academic teaching hospitals				
1	Beds:>1000 Specialists: 180 Personnel: 3800	Surgeons: 1 (I&O) Radiologists: 0	Consults, Vascular meeting*	Prints.
2	Beds: > 500 Specialists: 200 Personnel: 3000	Surgeons: 1 (I) Radiologists: 3 (I&O)	Bypass intervention*, Vascular meeting*, Cooperative scan assessment by 2 radiologists	Prints, screen (2D/3D in special cases)
3	Beds: >400 Specialists: +/- 100 Personnel: > 2000	Surgeons: 0 Radiologists: 1 (I&O) Technicians: 3 (I&O)	Patient scanning, scan manipulation (technicians 1,2), Scan assessment * (radiologist) Demonstration software ***(technician 3)	Prints, screen (2D)
4	Beds: > 580 Specialists: 150 Personnel: > 2600	Surgeons: 0 Radiologists: 1 (I&O)	Endovascular interventions*	Prints, screen (2D&3D)
5	Beds: >530 Specialists: n/a (Medical staff:>250) Personnel: > 2500	Surgeons: 1 (I&O) Radiologists: 0	Vascular meeting*	Prints, screen (2D)
6	Beds: >600 Specialists: n/a Personnel: >3000	Surgeons: 1 (I&O) Radiologists: 0 Technicians: 1 (I&O)	Consults*, Demonstration software***	Prints
7	Beds: >900 Specialists: n/a (Medical staff: >150) Personnel:>2300	Surgeons: 1 (I&O) Radiologists: 0	Cooperative scan assessment surgeon & radiologist *	Prints, screen (2D & advanced reconstructions)
Academic hospitals				
1	Beds: >1000 Specialists: n/a Personnel: >7700	Surgeons: 2 (I&O) Radiologists: 1 (O)	Endovascular intervention * Consults*, Vascular meeting*	Prints, screen (2D)
2	Beds: > 1000 Specialists: n/a Personnel: > 6000	Surgeons: 0 Radiologists: 1 (I)	Short tour radiology scan assessment rooms	Screen (2D)

Table 2. Brief overview of the nine hospitals and the activities and personnel observed. *Activity involves scan assessment, **All observation includes asking interview questions to the main participant * Demonstrations of software includes asking questions to demonstrating participant.**

5.2.2. Vascular disorders and treatments

The two main vascular problems are stenosis and aneurysms. It is not the purpose to give a complete medical overview, but to give the reader an understanding of these disorders and treatments. More elaborate information suited to non-medically trained interested parties can be found at for example the Vascular Diseases Foundation (2004) and Dutch Heart Foundation (Nederlandse Hartstichting, 2004) websites.

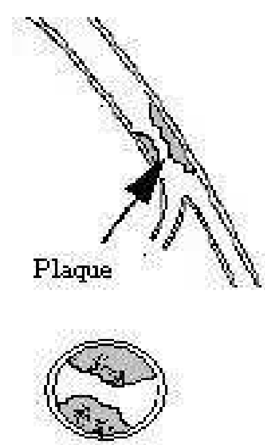


Fig. 17: Stenosis

Stenosis

Stenosis is a blockage of a vessel (fig. 17) due to the build up of plaque, consisting of amongst other substances fats and calcium, that stick to the vessel wall. The artery diameter decreases, obstructing blood flow and the plaque and blood clots (thrombus) sometimes completely block a vessel. The resulting diminished blood flow can e.g. lead to a stroke, or can result in a heart attack if vessels leading to the heart are affected.

Aneurysms

An aneurysm is a ballooning (widening, dilatation) of a part of a vessel. The VRE research currently focuses on disorders in the abdominal aorta, a specific disorder of Abdominal aortic aneurysms (AAA) is therefore discussed here. An AAA is a dilatation in the abdominal aorta (see fig. 18), that can in some cases lead to rupture of the aorta. A rupture of an aneurysm, especially in aortic vessels, can quite possibly be fatal for the patient and require emergency surgery. If an aneurysm is detected before it has ruptured, depending on its size and changes therein an intervention can be undertaken. Since unruptured aneurysms can go without symptoms noticeable to the patient, aneurysms are not always detected in time. The larger the diameter of an aneurysm gets, the greater the risk of such a rupture of the vessel. A current guideline for aneurysms states that when the aneurysm is larger than 4 cm in diameter or grows by more than 0,5 cm per year an intervention should be carried out if at all possible (El-Khalili, 1999; also reported by participants in this study).

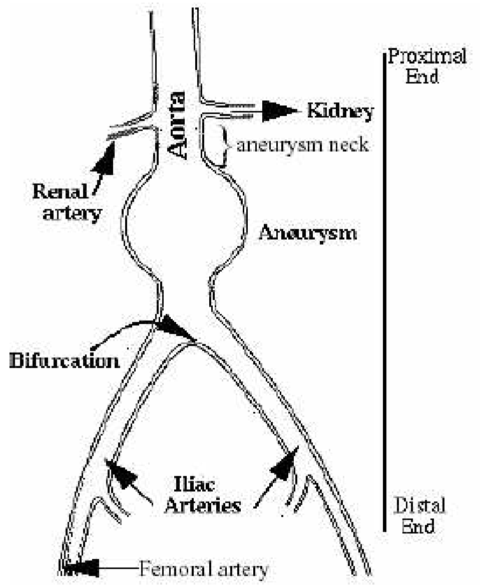


Fig. 18: Abdominal Aortic Aneurysm (El-Khalili, 1999)

Treatment

An array of treatments for vascular disorders exist. Medication can be prescribed to e.g. lower blood pressure (high blood pressure is for example a risk factor for rupture of aneurysms) or lower cholesterol levels and blood thinning agents to combat deposits of substances on artery walls. Lifestyle changes such as stopping with smoking, changing to a healthier diet and increase in exercise levels are always recommended to patients and are sometimes enough to combat minor vascular problems. If lifestyle changes and/or medication alone are not enough to combat the patient's disorder, an intervention may be performed. Some common examples are bypass placement, balloon angioplasty and endovascular stenting. These interventions can be performed by using open surgery or can be performed by endovascular treatment. Below, the most common procedures are discussed. It has to be noted that more interventional treatments exists, aneurysms can for example be clipped with small clips to decrease their size or filled with plugs or coils.

Open surgery bypass

Bypass surgery aims to reroute blood flow around the blocked or weakened part of an artery by placing an artery graft. Blood flow is diverted and the graft takes over the function of the faulty vessel length (fig. 19). The graft can either be an artificial stent-graft, or a patient's artery from elsewhere in the body. Bypasses are placed in open surgery via an abdominal incision by surgeons. For bypass surgery the patient has to be put under general anaesthesia.

Fig. 19: Bypass (adapted from www.gezondheid.be).

Minimally invasive ("keyhole") surgery
Both aneurysms and partially blocked arteries can in some cases also be treated using endovascular surgery. Such keyhole surgery is relatively new and has been in great interest and development since the 1980's (El-Khalili, 1999). While the VRE now focuses on bypass treatments, some of the interviewed surgeons expressed much more interest in applications helping them in endovascular procedures.

Endovascular surgery involves gaining access to an artery through a small incision and then guiding a wire with e.g. a stent to the affected artery. Imaging techniques are used to guide the procedure. Keyhole surgery is less invasive and risky than open surgery and allows the patient to resume regular activities sooner. Both surgeons and interventional radiologists can perform endovascular procedures.

A common endovascular procedure is percutaneous transluminal angioplasty (PTA) or balloon angioplasty (fig. 20). With balloon angioplasty a deflated "balloon" is guided to the stenosis that is to be treated. When the affected area is reached, the balloon is inflated, pushing the plaque to the artery walls and opening up the artery. The balloon is then again deflated and removed from the body. This procedure can be combined with stent placement; a mesh stent is then placed on the balloon and is left behind to keep the vessel open.

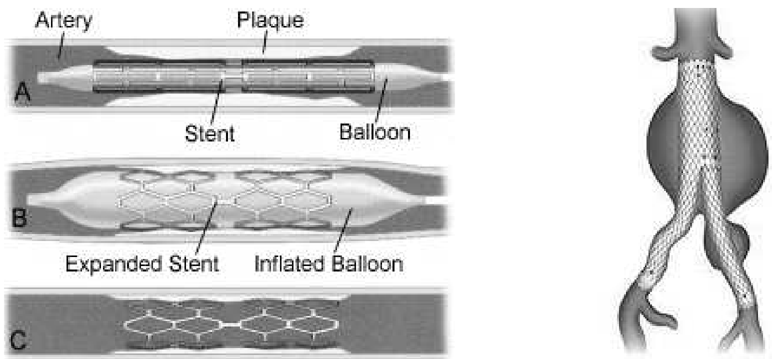


Fig. 20: Left: Balloon angioplasty and stent placement to counter stenosis (www.tmc.edu).

Stents can also be used to repair aneurysms, then fabric covered stent-grafts are used. Stent-graft repair of an abdominal aortic aneurysm involves making an incision in the groin and with a guide wire lead a tightly-packaged stent-graft to the aneurysm. When the stent-graft is in place, the covering of the stent is withdrawn, the stent expands and is fastened into the artery wall with hooks on the stent. Stents vary in diameter and length and special stent-grafts also exist especially for e.g. extension beyond the bifurcation into the arteries of the legs, see right image in fig. 20. Endovascular interventions are not carried out at all visited hospitals, especially smaller hospitals refer patients to academic medical centres for such procedures. More specialized information on endovascular procedures for abdominal aortic aneurysms can be found in e.g. Cuypers (2001).

5.2.3. Diagnosis and treatment planning trajectory

The first step in this study has been to determine the current medical trajectory of a patient with vascular problems, who chooses to get diagnosed and treated in a Dutch hospital starting with a visit to a vascular department. This trajectory has been studied to get more insight on how diagnosis and treatment planning for vascular disorders are currently conducted and what goals, tasks and decisions constitute these processes. This section will elaborate on this trajectory as it has been reported by the interviewees. Emphasis here is on the processes and tasks involving scans, not on medical issues.

Fig. 21 presents a simplified diagram of the main processes in vascular diagnosis and treatment planning. This diagram gives a simplified version of the trajectory for non-acute patients, acute patients may need immediate surgical repair and can be rushed the operating theatre and thus logically are not treated following the regular trajectory. A more elaborate diagram, broken down into 3 parts, can be found in the appendix 3 and should be examined for a more complete overview, including actors and information (e.g. forms, documents) that are the result of tasks and are exchanged between actors.

For reasons of clarity, some details and exceptions may have been left out of the diagrams. The elaborate diagram ends with the proposal for an intervention, but the complete process can be repeated when monitoring proves an intervention to be unsuccessful. It should also be kept it mind that after every test and scan a processing, interpretation and discussion with the patient can follow. These are not modelled, because the focus in this study is not on the information and consent processes involving the patient's non-clinical properties. Tests could have been carried out by other specialists before the patient reaches the vascular department, this information is then re-used if possible. Non-formal collaboration besides the vascular meeting are also common.

The representations of these processes are fairly high-level and do not yet study the individual planning and diagnosis tasks, but does study the process involving these tasks. Below the trajectory and its tasks are briefly introduced.

Consult

A typical vascular patient trajectory within a Dutch hospital starts with a visit to a (vascular) surgeon, or sometimes an assistant or trainee. Patients are referred to the vascular department by either 'eerste lijn' (primary health care, e.g. general practitioner) or other 'tweede lijn' (secondary health care) specialists with the request to confirm if a patient is suffering from an (aortic) vascular disorder.

A first consult with a patient him/herself is about 10-20 minutes long and entails anamnesis: diagnosis, identification of contra-indications of certain treatments, explaining of actions that can be taken to the patient and trying to motivate the patient to minimize risk factors for vascular disorders. In '1 or 2 minutes' (reported by first interviewee) the vascular surgeon reaches a preliminary diagnosis by asking a number of standard rule-based questions and possibly a physical examination (Note: An indication of the questions can be found in the Dutch standard procedures offered to general practitioners at Kaiser et al., 2003). After determination of the probable diagnosis, time is spent on contra-indications of possible interventions in a similar rule-based fashion. If contraindications to treatment are present, other tests are usually not performed during the period in which treatments are determined to be risky. For example only after 6 months after a coronary myocardial infarction (heart attack) the patient can be operated upon. Tests and treatments are explained to the patient as well, while additionally motivating the patient to minimize risk factors: for example to stop smoking and lower cholesterol-levels.

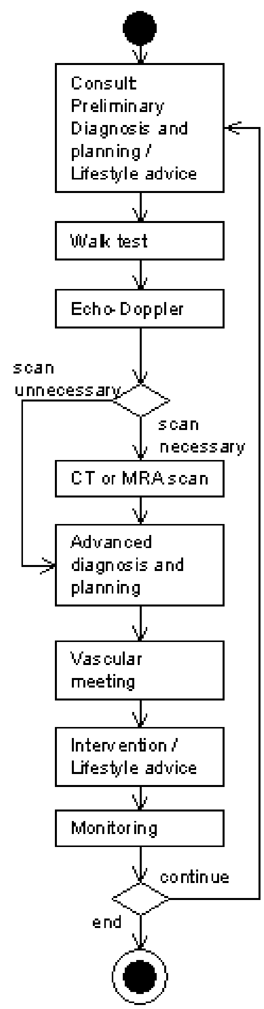


Fig. 21: Simplified trajectory vascular diagnosis and treatment planning

Examinations
Depending on the preliminary diagnosis additional examinations are conducted. This can be a *'walk test'* (strain endurance test on a treadmill) in which blood pressure are measured before and after physical exertion, resulting in time/pressure curves from which the expert can decide if a vain is obstructed, where the obstruction is and grade how severe the reduction in function is.

After the 'walk test' an *echo-Doppler (or Duplex) examination* can be decided upon. With this examination the vain can be pictured and the location of the vascular disorder can be determined. Echographies of aortic arteries are usually not the responsibility of the radiology department and are conducted by vascular technicians at the vascular department. A report of the examination is then handed to the treating surgeon.

Scanning
After assessment of diagnostic physical examinations, a walk test and/or Echo-Doppler examinations the vascular surgeon can decide to order a diagnostic MR or CT scan. These are conducted at the radiology department of a hospital.

A minimum of scans is made to gather all standard data. E.g. usually only one CT scan will be made, this to minimize the inconvenience and risk of using a contrast agent to the patient. The raw data of the medical scan and the subsequent post-processing is whenever possible performed by a technician to unburden the radiologists. The analysis of the images is the responsibility of the radiologist. Adjustments to the diagnosis and treatment plan are discussed with the vascular surgeon or returned in a report. More information on scan use can be found in the scan use section (5.3) of this thesis.

Planning an intervention
In short, planning an endovascular interventions or open bypass surgery after determining such treatment is necessary, requires information on what part of the patient's vessels should be treated what type of prosthesis would be most appropriate, what dimensions an prosthesis should have and where it can be attached to the vessel. This information is gained mainly from scan images.

It has to be noted that planning for bypass surgery is not finalized before an operation, with the operation just being a of a predetermined plan. Measurements to determine treatment is necessary are of course carried out before an operation, but decisions of e.g. length and diameter of artificial prosthesis' are often decided when the patient is open on the operating table, fitting and cutting the artificial vessel during the operation. This seems efficient and effective to participants, making them possibly less interested in exacter planning of e.g. lengths and diameters when a particular stent type has been chosen. A 3D visualization of the vessel seems less interesting when a patient has to be operated upon and the surgeon thus will see the vessel 'in real 3D' anyway, this might be a reason why participants showed more interest in support for endovascular interventions where stent fitting is a more pre-planned process and vessels cannot be physically examined from close range.

Vascular meeting
Most of the planned surgical and endovascular interventions are discussed with colleagues before implementation. The indications for diagnosis and treatment are then discussed in the weekly 'vascular meeting', which results in an advice for treatment. The irreversible decisions, save acute cases, are only made during these meetings. (Vascular) surgeons, (interventional) radiologists, surgeons in training, assistants and possibly other specialists discuss all the medical data of the patient available and decide on the final diagnosis and treatment plan.

In most of the observed meetings, the medical images are presented 2D on a lightbox or in rarer cases projected by beamer(s). Clinical textual data are either presented in hardcopy and read out loud by attendees or projected.

Interventional treatment
Depending on the nature of the planned intervention, the radiological or vascular department has to fulfil the request to perform an intervention. Interventional radiologists can carry out endovascular treatments; vascular surgeons perform open bypass surgery. Vascular surgeons do however also perform endovascular treatments and are very occupied with the recent developments in this field. After discharge from the hospital or a policlinical (outpatient) treatment the patient has regular consults and check-ups that become more infrequent with time, but can be lifelong.

Monitoring
After treatment, patients are depending on the treatments cared for at the intensive care unit and then at the vascular ward. Data that is monitored during this time depends on the particular treatment. After discharge from the ward, patients return to e.g. standard 'wound consults' and vascular consults to discuss results of treatment and possible further actions. For each treatment a particular schedule of return visits exists, varying from an endpoint at e.g. 3 months after treatment to lifelong monitoring, including possible scans to check vessels and possibly placed prosthetics.

Information used
Reports are used for information exchange and 'closure' of tasks to record examination results. Studying these reports has yielded very useful information about the data gathered in examinations and is used and needed in diagnosis and treatment planning.

The intake results in *anamnesis information*: the patient's story, his/her complaints and history.

The first *physical examination* results in outward observable artery data, and heart and lung function are the basis for (contra-)indications for treatments.

Walk test results in numerical data and curve graphs of distances a patient can walk before experiencing problems that can be interpreted afterwards to assess the seriousness of the patient's complaints.

Echo Doppler tests results in online 2D pictures, which cannot be easily interpreted after the examination. The examination visualizes the artery. Numerical data is presented on the velocity of the bloodflow. With this information the amount of obstruction/widening of the artery can be determined. Measurements of for example diameter are done by hand.

Scan data is used as an additional information source to verify preliminary diagnosis and for more extensive information, measurements.

The focus in this case study will be on scan assessment. While the above information sources have been studied, these fall outside the scope of this thesis and will not be further discussed. Measurements are described later in this results section.

5.2.4. Notes on trajectory
A number of observations that stood out in the found trajectory are highlighted in short below.

First of all, the studied tasks of diagnosis and planning are not as explicitly separated in the trajectory as expected beforehand. Preliminary diagnosis and treatment planning takes only a couple of minutes and is directly linked to possible treatments. Tests and possible scans are used for verification of the preliminary diagnosis and for more advanced treatment planning (e.g. for measurements). Before scan assessment, preliminary treatment plans were made 'in seconds' after hearing the patient's complaints during the first consult and some short physical examinations. The tasks following the initial consultation were mostly used to verify this first diagnosis.

This also entails that making scans of the patient is only performed when this is necessary for further verification and treatment planning. Only the reconstructions that are perceived as actually needed are made. It was surprising to find that when scans have been made, the actual time spent on individual assessment of scans was considerably less than originally assumed by the development team. Experts could assess 2D slices within minutes and often did so in between consults, phone-calls, conversations with colleagues and other tasks. No fixed times are scheduled for scan assessment. Extra subtasks, such as making a 3D reconstruction of scan data to gain extra insight in the geometry of complex vascular structures, are usually only performed when necessary, e.g. if reviewing of 2D slices has not yielded enough information. Efficiency is crucial in current medical practice; time is precious for the expensive experts.

Cooperation and consultation of colleagues is highly valued and almost all interventions are carried out only after informal and formal consultations and ‘vascular meetings’. All scan prints from selected orientations are discussed with colleagues during informal and formal meetings. Conclusions on final diagnosis and decisions on treatment were mostly drawn collectively by radiologists and surgeons during vascular meetings - although the patient has the last say in his/her treatment. Chances for aiding both formal and informal consultation are apparent and cooperation should be supported by systems that aim to facilitate more efficient diagnosis and treatment planning.

5.2.5. Human actors

A number of human actors (i.e. human agents, systems) are currently involved in diagnosis and treatment planning for vascular disorders. Below focus will be on those actors who are involved in the use of scan images in this processes.

In development of the VRE, originally two user groups were considered: radiologists and surgeons. Unexpected at the start of this study, another group was identified that plays a considerable role in the use of scans in diagnosis and treatment planning: radiology technicians. These technicians are responsible for the physical scanning of patients. However, they do not only simply gather scan data, but also prepare scan images for e.g. radiologists and surgeons. Prints or digital images systems used are most extensively handled by technicians. They try and ensure that e.g. radiologists and surgeons can efficiently assess scan images. This extra potential user group has to be included in design and implementation considerations if the VRE were to be used in a clinical context.

Depending on the complaints and location of the vascular disorder, other specialists may be involved in diagnosis and treatment planning and scan discussions. These specialists include internists, cardiologists and neurologists. Even while scans may be discussed together with these specialists, these involvements mainly seem to involve consultations. Actual scan manipulations are performed by the three groups outlined above, these extra groups of specialists are not discussed in this thesis as potential end users.

It is also acknowledged that system support staff will be needed in implementation and maintenance in a hospital. Additional medical specialists and support staff are however not extensively discussed in this exploratory study. Below each of the three prospect groups of end-users (radiologists, surgeons and technicians) of scan processing and diagnosis/planning systems for vascular disorders are discussed.

Vascular surgeons

Vascular surgeons are the primary group of medical staff dealing with patients with vascular disorders. They are the primary group to consider when developing diagnosis and treatment planning systems for vascular interventions. Ultimately vascular surgeons are responsible for the vascular patient's whole trajectory from intake consult to patient discharge. They perform patient consults, diagnose patients, plan treatments, perform treatments and monitor patients. This in contrast with only performing an intervention itself and then returning the patient to the treating physician. Additionally, most of the surgeons who participated in this study, are actually surgeons with a specialization in vascular disorders and perform all kinds of surgical procedures.

Responsibilities

Vascular surgeons are not only responsible for patient contact and treatments. Apart from clinical activities, they also both advise and consult colleagues and occasionally perform management functions. Many vascular surgeons are also heavily involved in educating medical students and surgeons in training. Some are also involved in research projects and development of new treatment and consult protocols. Protocols, norms and rules regulate all of their activities

Pressures

Every working day surgeons offer advice and make decisions that can quite severely impact patient lives. They have to deal with emotional reactions of patients while trying to treat them as effectively and efficiently as possible. Each working day, surgeons have to deal with time pressures in respect to the numbers of patients that have to be treated or receive a consult. All work needs to be accomplished as efficiently as possible. One surgeon who was followed during his office hours saw about 35 patients during that time, including consults and occasional scan assessments. Added direct time pressure occurs in case of acute patients; all other work needs to be abandoned immediately if an acute patient is brought in and the surgeon is the only attending specialist present. Apart from emergencies, further interruptions occur with beepers going off demanding immediate return calls, patient calls to the department and colleagues asking for informal advice.

Domain expertise

Vascular surgeons are domain experts. They have had years of extensive medical training and have to keep up with current medical developments as well. Surgeons needs to match diagnosis criteria with possible disorders patient may have. They have to interpret information acquired through listening to patient's stories and history, through physical examinations and by using medical imaging techniques. Contra-indications of treatments and risk factors have to be considered in finding the optimal treatment for a patient's condition.

Computer expertise

Relying on their expertise and experience, surgeons might feel they do not actually need new technology to effectively treat their patients. A certain scepsis exists in the participants towards claims of usefulness of new systems under development. Most are however interested in new technology and some participants themselves are involved in technological research projects that possibly could be relevant to the VRE research, such as development of expert system aiming to support choice between endovascular stents. The participating surgeons were usually unaccustomed to working with VR systems, but all had some experience in using advanced equipment and most did use PC's for work.

(Interventional) radiologists

Radiologists are the principal group to consider when developing systems that aim to support assessment of scan images. Radiologists are trained in the use of medical imaging techniques for diagnostic and therapeutic uses and responsible for assessment of these images.

Responsibilities
In the vascular trajectory radiologists are responsible for the quality and assessment of scan images as requested by vascular surgeons. They can both offer advice on treatment to surgeons, or in the case of interventional radiologists, actively plan and perform key-hole (endovascular) interventions, such as placement of endovascular stent placements.

Like surgeons, radiologists can be also responsible for teaching and supervising co-assistants (interns), can be responsible for possible management functions and may undertake research activities.

Pressures
Radiologists too have to deal with time pressures; large numbers of patient imaging procedures have to be performed and assessed. Just like surgeons, radiologists' decisions can seriously impact other people's live and they have the responsibility to carry out their work accurately. Work can be interrupted quite often by calls or colleagues walking in.

Domain expertise
Radiologists are physicians, domain experts on imaging techniques, but have also been trained as physicians and can offer advice on diagnosis and treatment of patients. Interventional radiologists have additional knowledge on non-surgical interventions, such as endovascular stent placement.

Computer expertise
Radiologists are used to working with complex imaging equipment and (PC) workstations. None of the participating radiologists had experience with VR, however they are used to working with 3D visualization software. Some of these can be considered 'fuzzy boundary' like-VR systems offering e.g. 3D visualization of anatomic structures or walkthroughs of structures such as the colon or arteries.

Radiology technicians
Radiology technicians are deployed with the actual data gathering and patient contact during scanning. They process scan images and present radiologists with a set of images to assess.

Responsibilities
The tasks of radiology technicians include direct patient contact, actual acquiring of scan data, checking scan quality and post-processing scan data into images to be delivered to the radiologist. Technicians prepare the pictures and orientations that are presented to the radiologist to assess. They choose visualizations and viewpoints for printing and/or saving in digital patient file, operate scanning equipment, printers and post-processing workstations.

The set of images delivered usually follows standards for the requested scan and suspected pathology, but a radiologist may request additional images or manipulate data himself. Technicians are responsible for the quality of the acquired images and formally have no say in the interpretation of the images. If the technician is very experienced, reviewing datasets on workstations is sometimes performed in cooperation. Ultimately the radiologists is responsible for medical image assessment. Experienced technicians may additionally mentor technicians in training during their work.

Pressures
Each scanning procedure and post-processing of the scan data has a predestined timeslot in which work should be completed. Dealing with patients, instructing trainees and interruptions due to calls and colleagues walking in during work.

Domain expertise
Radiology technicians receive a specialist college-level ("HBO") education. They have to have knowledge on and technical properties of the various imaging modalities and their medical applications. They have a basic medical training with knowledge of e.g. human anatomy and pathology (disorders and diseases) and techniques for handling patients. For scanning procedures and image processing the technician needs to know what anatomy to scan, what abnormalities to look for and what settings to optimally show anatomy and pathology.

Computer expertise
Radiology technicians are used to working with scanning equipment, workstations and post-processing software. Technicians usually specialize in a particular kind of imaging modality (e.g. MR or CT). At some hospitals the available technicians 'rotate' and work with all different imaging techniques used at that hospital. Technicians are often provided with specialist training for working with a specific imaging modality and might not be able to work with all systems available in their department. The observed technicians did not have any experience with VR systems, but did occasionally work with 3D capabilities of software on workstations.

Specialist (vascular) treatment planning systems will probably not be used by technicians, but it is not implausible they may have to provide the input to future systems, including the VRE, if actually implemented and manipulate scan images in visualization systems. They have to be considered in system development

5.3. Scan use in diagnosis and treatment planning

When after echography during the echo-Doppler test more information is needed on the patient's vascular problems, MR(A), CT(A), and angiography are the main imaging modalities used in the further diagnosis, planning and treatment process. These modalities can be used for example to examine the extent of stenosis or to detect an aneurysm. Angiography can also used to guide endovascular procedures, while CT and MR imaging are used in diagnosis and planning e.g. to gain precise measurements.

This section first introduces these modalities, then the various available reconstruction projections, from slices to 3D imaging, are discussed. The use of these projections at the visited hospital departments and the found issues surrounding scan use are discussed afterwards.

Angiography (or rather X-ray angiography) visualizes vessels using X-rays (fig. 22). The various tissues of the human body absorb the X-rays in different gradations, producing a distinct appearance for each type of tissue on X-ray images. A contrast agent has to be injected to make blood visible on the X-ray. This can be achieved by inserting a catheter, leading it up to the desired location and then injecting contrast fluid. Individual X-ray images can be stored on e.g. video or digitalized and played as a movie to view blood flow in time.

Fig. 22: Angiography of part of the abdominal aorta (Rochat, 2002).

Computed Tomography Angiography (CTA) also uses X-rays and contrast agent injection to visualize arterial blood flow. In CT scanning this is combined with computerized image analysis. A CT scanner (fig. 23) takes X-rays of a point of interest in the body from various positions around the body. Scanner and patient are moved producing a spiral series of X-rays. These images are then combined and processed, allowing for various 2D and 3D reconstructions. In CT *angiography*, various images are combined and processed to picture blood flow during a period of time to visualize the whole of the desired artery length.

Fig. 23: CT scanner (www.philips.com).

MR (magnetic resonance) images are very similar in appearance to CT scans. However, MRA (Magnetic Resonance Angiography) works using a magnetic field and radio waves instead of X-rays. The MR scanner's magnetic field aligns hydrogen molecules, the scanner emits and picks up radio waves; each type of tissue will return a distinct signal due to the alignment of the amount of hydrogen in that tissue. Like CT angiography, MR angiography is especially aimed at picturing blood flow.

The choice for MR or CT depends on a number of factors. Certain properties of MRA scanning prevent the visibility of calcium on scans. Assessment of calcifications in vessels and assessing amounts of thrombus and plaque are however very useful in diagnosis of aortic problems, therefore CT scanning is often preferred for e.g. diagnostic images of aneurysms. Contra-indications for CT scanning thus includes patient's allergies to contrast agents and problems associated with their intra-venous injection. MR scanning is also generally considered less hazardous for the patient than CT scanning, since no X-rays are utilized. MR scanning is due to the use of a magnetic field however not suitable for patients with e.g. metal prosthesis's, pacemakers or other metal/magnetic objects, since the magnetic field might move or negatively affect function of such items. Participants reported that availability of scanning equipment at the hospital might play a role as well in the choice for a certain imaging modality.

Elaborate information on MR(A) imaging can be found in e.g. Nagel et al., (2002) and MR-TIP (2004). For more information on use of CT in monitoring vascular abnormalities, specifically after endovascular repair of aneurysms, see Wever (1999).

Projections

Various image reconstruction techniques are used with MR and CT angiography: axial slices (cross-sectional), maximum intensity projection (MIP), multiplanar reconstruction (MPR), shaded-surface display, volume rendering and curved reconstructions.

Traditionally, scan images are viewed as 2D slices ("coupes") from sagittal (cutting from front to back), coronal (bisecting the body into front and back parts) and axial (cutting the body into top and bottom parts perpendicular to the long axis of the body) orientations. Prints of 2D images, or 2D images of 3D reconstructions, are often viewed on lightboxes (fig. 24) with multiple prints showing a variety of orientations in varying numbers of slices per print. Slices can also be viewed as a movie in so-called Cine-mode; offering a faster overview of structures by providing an automatic walk-through of a stack of slices.

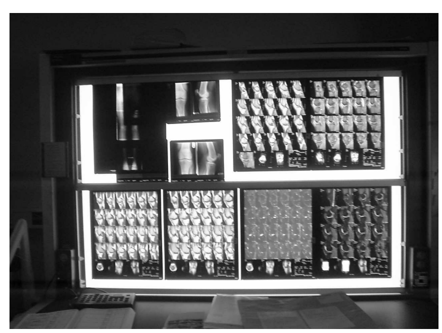


Fig. 24: Scan prints on a light box

Maximum intensity projection (MIP) (fig. 25) takes advantage of the intensity of vascular structures. A MIP reconstruction projects the point with the highest intensity in a 3D dataset into a 2D image. Encountered blood vessels result in high intensity dots, the MIP connects these in three dimensions, providing an angiogram that can be viewed from any projection. (MR-TIP, 2004)

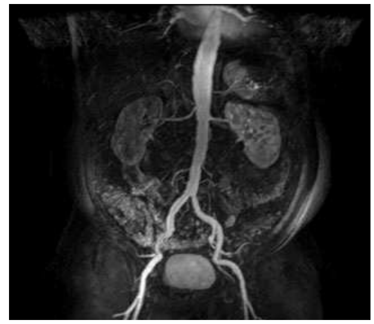


Fig. 25: MIP of a contrast MRA of the abdominal aorta (Nagel et al., 2002).

MPR (multiplanar reconstruction) post processing reformats 3D data set into 2D slices of arbitrary thickness at any angle. Thus producing 2D images from any orientation the user would like.

Curved reconstructions can be of help in exploring artery condition. Arteries are not straight lines and may contain curves that are not completely visible on one regular 2D slices, which follows a non-curved plane. In curved reconstructions the image plane follows the trajectory of the artery. The whole desired length of the artery, including all curves, can then be viewed on one 2D image. Virtual endoscopy software (fig. 26), can offer a fly-through mode following this same trajectory, in this way arteries can be explored from the inside.

Volume rendering and Shaded Surface Display (SSD) offer three-dimensional images (both referred to as 3D renderings in this thesis), various software options enable users to manipulate 3D images using tools such as zoom, pan, rotation. To not have view obscured by other objects than the object of interest, organs can be cut off the image, or translucency of volumes can be adjusted.

Fig. 26: Virtual endoscopy of the aorta (Nagel et al., 2002).

How the above projections are used by the participants in this study and the main encountered issues in assessment of scan images are highlighted below. Links to the functionality of the VRE will already be made here as well. Since the focus of the VRE lies on 3D visualization and planning for abdominal vascular problems, the emphasis here will lie on how 3D renderings are now used, after which the information and measurements needed to plan treatments are discussed.

Prints vs. workstation scan assessment

Even though assessments of 2D or MPR/MIP reconstructions are more often conducted on workstations, prints are still used often. Although sharing of prints is less convenient than digital files, prints can be carried around, assessed on wherever a light box is present and discussed with patients during consultations. Access and efficiency are important. Using prints seems convenient for surgeons, because essential views and orientations are provided, ready for assessment. However, scan assessment on workstations (fig. 27) is commonplace for radiologists and becoming more common for surgeons. Working with datasets themselves, they do not need to rely on technician's skills to choose the right projections and actually show them the images that are important for the particular patient. A more complete overview of scans can be gained

Fig. 27: Scan assessment using workstations.

using systems themselves by e.g. using the CineMode to view of a stack of slices or possibly reviewing a whole 360 degrees view of a MIP reconstruction, instead of printed 2D images from a limited number of orientations.

On both screen assessment and inspection on light boxes multiple images are often viewed simultaneously, either by hanging various prints on a light box together, viewing various orientations on a single screen or by using multiple screens. This should be supported in a new tool aiding scan assessment.

Use of 3D

Currently, no intervention is carried out by evaluation of 3D visualizations alone. Training of radiologists and surgeons still focuses on the assessment of 2D images of scan slices. The interviewees indicate that they mentally reconstruct the 2D images to 3D anatomical structures.

3D information is always used to gain MIP and MPR reconstructions. The scan data is always looked at in 3 orientations and thus ‘in 3D’. Often, technicians select a number of orientations and print these for clinicians to assess, offering them 3D information on 2D slices. 3D renderings are not always called for by clinicians. Assessment now mostly is conducted using these selected 2D prints from various orientations.

MR and CT scanning thus results in 3D data, and this data is used, but 3D renderings and 3D images might not be constructed. Even though 3D visualization is now available in most hospitals on workstations in radiology departments, it was found that the available systems are not always used for assessing vascular problems.

Combining 2D and 3D

3D is often just an extra and deemed unnecessary to gain sufficient information for diagnosis and treatment planning. If used, 3D visualizations are appreciated most to get an overview of intricate vascular structures or complex cases. The more difficult the case, the more complex post-processing is performed. 2D scans are then still used to review details and to perform measurements. A system such as the VRE would need to support the need for 2D visualization as well as advanced 3D functionality; convenient switching from one rendering to another would be appreciated. Presenting 2D and 3D rendering together provides for more efficient assessment.

Efficiency

One of the main reasons not to use 3D renderings is efficiency. Reportedly, time concerns prevent routine 3D segmentations. Processing of 3D reconstructions take considerable time
The processing speed is important. During demonstrations of software, or observations of regular scan processing; copious complaints about slow hard- and software were made.
Sometimes human effort and time are required as well, since completely automatic processing is not always available for the particular desired reconstruction. Software can automatically generate 3D pictures from scans, but to achieve a view. Software might recognize vascular structures, but locations of obstructions and such are often still manually indicated by regular (non-interventional) radiologists or technicians. additional manual input is often needed to outline organs and cutting them from view when needed.

Surgeons do not usually have a scan processing workstation at their work places at vascular surgery departments, thus 3D visualizations take even more time going to the radiology department and asking for other’s effort and time; adding to the inefficiency of using 3D in their tasks. Equipment might not always be available to them when needed; either the hospital might not even have the equipment, or someone else needs the equipment more. If the use of 3D reconstructions would be standard, surgeons expressed the wish to have all reconstructions available at their own department, at best in their own office.

Only when the added value of 3D rendering weighs against the time it takes to construct and interpret a 3D visualization this will be used. This is a good reason to think about a better performance and an efficient interface. Improvements in automated segmentation of anatomic structures in medical images thus seem very desirable as well.

The only time intensive use of 3D volume renderings has been observed during this study was actually in a radiological department's operating theatre (fig. 28). The program was used during endovascular procedures to review angiography images before and after the procedure.

According to the participating interventional radiologist this program was used because it was fast, efficient, easy to use, offering automatic segmentation, clear orientation and manipulation of the 3D object. The distinct advantage seemed to be the possibility to review the images together with other medical staff during the procedure and fast segmentation. However, even though the room with the system is now only steps away, having the 3D images available within the intervention room, like the 2D reconstructions already are would be easier and more efficient (fig. 28 shows only monitors in the extra room next to the intervention room, behind the glass however more monitors offering the 2D views are available) Now e.g. gloves now have to come off and the radiologist has to walk back and forth behind the glass a number of times to view the 3D images. The VRE currently cannot provide for such efficient use, nor does it provide for fast processing of larger datasets; vast improvements are necessary if the system is ever to be implemented in a fast paced clinical environment.

Reliability and trust

Reliability of 3D reconstructions is an issue as well. Participants seemed hesitant to make use of any new visualization. Most vascular surgeons for example have been educated in the past using CT catheterisations and contrast agent flow interpretation. Now they have to trust software to generate accurate 3D pictures of e.g. MRA slices and they themselves have to learn to interpret new forms of visualizations. Some are still a bit unsure when they only use MRA and need to check if the visualization used in planning was correct during interventions. Reliability of systems and reconstructions need to be demonstrated.

A number of participants report they feel that 3D rendering might obscure internal condition of organs. As one participant explained; when the complete scan dataset would be considered to be 'the truth', tricks and renderings at their best show this truth, but can just as well obscure it. If for example, an abnormality is present on the inside of a structure, a non-translucent 3D rendering from the outside of the structure renders this possible abnormality invisible. These abnormalities can however always be seen on slices (if their resolution and in-between distance are adequate). Judgements in 3D are therefore always checked using 2D slices. Translucency settings of 3D renderings, or new solutions need to be developed if 3D renderings are to be used more often in scan assessment.

A number of problems cannot be solved by offering new renderings of scan data. Clinicians have to be critical of scan data. CT/MR can provide for overestimations of pathology and stents may not always be identifiable. Low scan quality by e.g. scanning with wrong setting cannot be solved. Neither can discussions to ascertain if scan abnormalities are scan artefacts or actual physical abnormalities be prevented.

A number of participants state it's not the pictures that need to be treated, it's the patient's complaints. If for example the left leg shows a severe obstruction, but the patient suffers on the right leg, the patient is not helped by treating the left leg and the real reason for the complaints may go untreated. The difficulty lies in deciding when "enough is enough" in diagnostics. The possibilities in patient imaging are getting larger, but interpretation is key. It is hard to interpret new abnormalities that are becoming visible on scans; which have to be treated and which are not harmful to the patient is not always clear. Deciding when enough information is available on the patient and the diagnosis process can be finalized is getting harder with such new possibilities to image previously unnoticed abnormalities.

Actual added value of new types of 3D rendering and it's reliability need to be proven; a certain scepsis to their use will (perhaps only initially) remain.

Skill
Reliability of reconstructions depends a lot on technicians' skills. The systems have the functionality, but not everyone has the knowledge and skills to use those functionalities. Training is necessary for any new functionality that might be introduced and its value has to be demonstrated to clinicians who would use the new types of images. A sentiment expressed is that it is impossible to know about all new techniques and visualizations; one participating surgeon considers all manipulations to be relatively difficult ('you have to click all kind of buttons and stuff', it's better to let the experienced people show you how it's done').

If fast and reliable reconstructions were possible, some of the participants would like 3D reconstructions to be standard, others seem less keen since they can already treat patients successfully using only 2D slices.

Other 3D utilities
Specific 3D "VR" functionality is available at all visited hospitals. Standard software used to process scan data often offers features such as virtual endoscopy. The VRE also offers this functionality, offering users the possibility to move through vessels. However, reportedly this functionality is never actually used at these hospitals and seen as more of a gadget than a valuable diagnosis and planning tool. One participant does state Walkthroughs could be useful for smaller vascular structures, more than they could be useful for the viewing of larger vessels such as the abdominal aorta which is the focus of the VRE. Thus marketing this virtual world utility to radiologists and surgeons as a valuable new addition to their existing tools seems rather futile. Especially since the virtual endoscopy functionality offered by available post-processing software is already rather advanced and seems more easy to use in some respects than the functionality currently offered by the VRE. For example, usability issues of the VRE such as getting lost outside the artery are impossible by restriction of navigation through the artery to a line through the vessel's centre.

Stereoscopic imaging
Stereoscopic vision of 3D reconstructions is never used, nor available, in the visited clinical departments. Because they are viewed on a 2D desktop they are essentially 2D views of a 3D reconstruction. If actual 3D stereoscopic vision is ever to be used, the added value needs to be proven and has to weigh against the effort of for example using stereoscopic glasses.

Specialized vascular treatment planning systems used
Use of some specialized vascular analysis software has occasionally been observed during this study. Fig. 29 shows an example of such a vascular analysis software package. These software packages offer automatic measuring (although measurements and automatic vessel centreline detection often have to be checked by manual input) coupled to a curved planar reconstruction offering a view of the whole length of the vessel in one slice, resulting in a graph of vessel diameter. In fig. 29 for example, this graph shows a diameter enlargement and its location within the artery. Along with the view of the whole artery, a view is given from various orientations; 3D images are offered, along with e.g. coronal/saggital 2D slices and oblique coupes perpendicular to the vessel. Minimal and maximal diameters, area, lengths, angles can be measured. Widths and distances measured to certain landmarks (e.g. healthy vessel lengths, or the bifurcation) can be used to determine prosthesis size.

Some of such vessel image analysis systems also offer advice on the prosthesis that might be right for the patient (e.g. General Electric's AW 40) These were not observed to be used in the visited departments.

Fig. 29: Vascular analysis software (www.philips.com). Coupes with measurement lines are shown above left, a graph of vessel diameter and a curved reconstruction of the vessel are shown below. On the right 2D and 3D images of various orientations can be viewed.

These specialist systems are available, but most of the planning for vascular interventions was done with scan analysis software or scan prints, measuring diameters by hand on 2D scan coupes, and images were usually not analysed using these types of software. A need for more specialist support in stent planning has however been demonstrated by participant's own endeavours to development tools to gain this support. Two of the visited departments have been involved in the development of separate expert systems aiming to support the preoperative planning of Aortic Abdominal Aneurysm (AAA) endografts. It tries to assist medical practitioners with assessing which brand/type of a stent would be the best to treat a case of aneurysm. The system is based on the standard AAA endografting preoperative worksheets used in the departments and developed in close participation with prospect end-users, thus aiming to tailor the system to user's utility needs and fitting it within the context of use.

Measurements and manipulations

The VRE should also provide for a number of measurements and manipulations if the system is to be implemented in a clinical setting. A measurement tool's ease-of-use and precision are crucial. Some measurements which could be helpful according to some participants, are not carried out because they are not strictly necessary and take a lot of effort to obtain.

The main manipulations performed on scan images during assessment include:

- Axial, coronal and sagittal movement through stacks of 2D slice images (including automatic movement in Cine-mode)
- Changing axial, coronal and sagittal position in space.
- Changing orientation.
- Zooming.
- Adjusting of width and level to get optimal settings for viewing anatomy and disorder of interest.
- Adding of text to prints.
- Adjustment of transparency in 3D renderings

Measurements conducted in both 2D and 3D scan images include:

- Lengths (e.g. determining affected length of artery, and needed length of prosthesis, thickness of walls and e.g. calcifications)
- Diameters (e.g. determining gravity of calcifications, thrombus, blockage or size of aneurysm, and needed stent diameter)
- Areas
- Angles (curves in vessels make placing endoprosthesis difficult and can even prevent endovascular treatment).
- Volumes (while some interviewees claimed never to use this measurement, it can be used to assess size of aneurysms (Wever, 1999)
- Pressure/ Elasticity of vessel wall is measured in rare cases

Pressure and flow measurements are reportedly not derived from scans, although flow measurements can be performed on scan data (e.g. by assessing relative positions of contrast agent).

For abdominal vascular disorders a view of the aorta is needed that spans the vessels to the kidneys, to the iliac arteries to the limbs after the bifurcation of the aorta (as seen in fig. 25) Blood flow should always be unobstructed to ensure functioning of limbs and organs. It is thus very important that these are not blocked, or won't be blocked by e.g. placing an impermeable stent graft that is too long in the wrong place, bypassing one of these arteries. The VRE testing data set does not include this data, but since this relies on the dataset that has been gathered using the scanner, it can be argued that this is not a design flaw. However, in automatic measuring systems, the bifurcation and arteries to the kidneys have to be recognized and considered for correct automatic interpretation of the aorta trajectory and e.g. diameters.

Automatic measurement systems would be greatly appreciated by a number of participants. Currently, a number of problems exist in standardization of measurements. Surgeons often want to perform measurements themselves to verify that e.g. diameters have been measured at

correct place. Percentages of stenosis can be under or overestimated when measurements are conducted in a suboptimal position. If, for example, a vessel's 'healthy' diameter is measured on a position where plaque has already decreased the vessel's diameter and the percentage of blockage is then computed using a comparison of this 'healthy' vessel diameter,
with the minimum diameter of the area affected by stenosis, the seriousness of the stenosis can be underestimated and treatment postponed.

Performing 3D measurements is reported to be especially difficult in current tools. Not only the desired position of a measurement, but actually placing markers on these spots is still hard. It is difficult to determine whether a marker's position in 3D space is actually correct. The 3D images are essentially viewed in a 2D projection on the 2D computer screen, possibly giving the user a distorted sense of where measurement markers are actually placed. If the user does not carefully explore the volume's rendering from all sides, incorrect measurements might be carried out, optional automatic support could possibly help prevent this. Whether stereopsis would help users notice such positioning problems in these kinds of situations is unclear; little literature could be found on this particular problem.

5.3.1. Blood flow simulation and bypass planning

The VRE visualizes simulation of blood flow before and after a simulated intervention with placement of a stent prosthesis/ bypass. Now, detailed blood flow information is often not used in advanced treatment planning. A reduction in diameter and thus flow in vessels is assessed, but no numerical data on flow is usually considered after assessment of early tests. Pressure and flow measurements result from Duplex testing and physical examinations, not from using scan images. Quantitative blood flow measurements are not commonly used beyond for determining whether treatment is necessary.

Simulation in the VRE will only be used if its added value to the quality of a treatment and the efficiency is clear to the prospective user, which is not the case with its current functionality. Not specifically flow in the vessel as the VRE provides, but more specifically outflow towards organs (e.g. the kidneys) and extremities (arms, legs) is of interest. The simulation offered by the VRE also focuses on the intervention result in blood flow in an artery, right after treatment. Participants indicated that this direct result might be important, but that a reliable prediction of long-term results of an intervention in terms of re-stenosis and blood flow would be more valuable. Some findings suggest that blood flow simulation would only be used at all in an adapted form used in prediction of wall stress, vessel elasticity, calcification, and survival of a stent prosthesis over time. However, information about blood flow on specific points in the artery can provide insight in where in the vessel re-stenosis might occur, by viewing where 'turbulence' might occur with part of the blood slowing down or re-circulating in a particular area and potentially depositing plaque substances. It has to be noted that most participants overall were sceptical of use of blood flow simulation data. Verification of any simulation data using patient data is crucial for acceptance; trust has to be established and regulations fulfilled.

5.3.2. Social and organizational environment

Regulations, formal rules and hierarchical responsibilities govern the process of diagnosis and treatments. These rules will have to be taken into account for any system developed in this context. Use of new systems, such as the VRE, may have to be approved by medical governing bodies in some areas and protocols for image processing, accuracy and reliability of measurements will have to be studied for a system to be trusted and approved.

Collaboration

Both remote and collaboration on location should be considered. Collaboration between colleagues (e.g. fig. 30) is highly valued and should be supported in a system used in diagnosis and treatment planning. A common complaint observed in especially larger vascular meetings is that not all participants get equally good view of scan prints or images on screen. While one visited department solved this problem by using beamers, this solution will not necessarily do for VR systems. In a stereoscopic VR system multiple participants will have to be offered the same view of 3D objects of interest; currently not necessarily the case with the VRE.

Fig. 30: Collaboration during scan assessments on light boxes amidst other equipment.

Exchange of data between departments and hospitals could be supported as well. The Grid developments in relation to the VRE, as discussed in section 3.3 “The VRE”, might prove useful for this by providing a network between hospitals and distributed computational capabilities for imaging data processing

5.3.3. Physical environment

The physical environment of radiology and vascular departments can pose a number of restrictions on systems. Certainly, if systems are to be used in an operating theatre, system hardware needs to be sterile or packaged in a sterile way, but light and noise levels as well as physical space can play a role as well. Most of the facilities and rooms used for assessment of scans are dimly lit to facilitate the human eye in assessing scan prints and images on screen. A system to be developed for such use has to be usable in such a dimly lit environment. Possibly, hardware knobs and buttons have to be distinguishable without full lighting, although such specific measures did not seem to be necessary for use of current equipment and systems.

A system has to be usable in a noisy environment as well. The clinical environment is hectic and noisy. Large numbers of people are present in the departments, discussions are held, sudden emergencies occur, beepers go off and calls have to be made. One large radiology department used cubicles with workstations in a large room for post-processing, amongst other work. Such situations make using voice recognition, as the VRE does now, somewhat (if not rather) challenging. Some of the participants also dictate reports of their assessment of scans directly while looking at them and taking measurement; this makes work more efficient. A possible new system should not prevent this and voice recognition might thus not be the most viable option to use in such an environment.

When electronic patient files are used, often other documents and information are being viewed simultaneously on extra monitors and stacks of paper files are kept at hand. The used equipment in scan assessment are mainly large automatic light boxes, pens for pointing, rulers for measurements, voice recorders and telephones. People carrying in other equipment, (metal) carts with e.g. paper files or already present (scanning) tools and equipment should not disrupt a system. Magnetic tracking as is implemented in the VRE right now, which is so sensitive that any metal material in the room may disrupt the system, is not suitable in such an environment. This demands either another tracking solution or a vastly improved magnetic tracking system.

Some hospitals have rooms especially for group assessment of scans, such as vascular meetings, where beamers are used, with a staff member operating (a) workstation(s). Portability of a system might prove useful, as dedicated space may not be available. A large VR system and screen possibly standing next to, or replacing, large light boxes is certainly possible in larger departments, where large rooms are reserved for e.g. scan assessment (as in fig. 31). However, especially in surgeons and radiologists own offices or in operating theatres, usually no dedicated space is available for an extra, large screen. Additional physical space is needed for the user to manoeuvre in. A smaller or desktop solution would be more suitable in such situations.

5.3.4. Technological environment

Apart from the systems that are explicitly meant for scan assessment and occasional specialized vascular treatment planning, a wide array of technology, software, tools and equipment is already in use in radiology and surgery departments. Possible input from and output to these systems needs to be considered when developing a new system that is to be implemented in that environment. Mostly workstations using Philips EasyVision and Siemens Leonardo systems were observed to be used in scan assessment; the VRE will need to be able to exchange data with such systems and also offer value over their utility and usability in vascular diagnosis and treatment planning.

Effective treatment and treatment planning is a principal goal, but a lack of efficiency and processing speed is often reported as the main problem in use of systems that aim to increase this effectiveness. Less is more; functionality that is not needed, but still included in a program is not appreciated and long processing times are unacceptable. Both ease of use and effective training are paramount.

Integration and compatibility with other systems used is important; some available systems were not used, because they could not exchange data with other systems and printers. Output of scanning equipments/systems is currently usually in the 'Dicom' file exchange format. Any system to be used in scan assessment should be able to deal with such in- and output. no standard set of resolution and size of scan images can be expected though; datasets depend on a number of factors, including equipment settings and anatomy scanned.

Prints can still be required as well when a computerized system is in use. Even though scan assessment will probably be done without hardcopy prints in the near future, currently, the need for a system to have the option to print images cannot be discarded. Most of the visited hospitals were in varying stages of transferring to use of electronic patient files and Picture Archive Communications Systems (PACS). Not all patient data was available electronically yet or electronically exchangeable between departments. Clinicians did occasionally have access to textual patient data, reports and textual test results, but digital scan data was usually not accessible. If a PACS was used, it did not accept 3D and MIP datasets. With current PACS data available at the hospitals reconstructions and new renderings cannot always be made. One of the great problems of digitalisation of the radiology department is the need for very large digital storage capacity. It cannot be expected that all data needed as input for a 3D reconstruction will be saved. Currently, raw scan data is saved for a limited time, sometimes now for only hours to days until workstation storage space runs out.

It has to be kept in mind that many systems that aim to support scan assessment are already available from established companies. New systems are regularly tried out, but many system are acquired as a bundle together with new scanners. This can make entering the medical market with a commercial system rather difficult. Problems of systems already used now in the treatment process have to be overcome for a system to impress and be an asset to users. Efficiency, time and training needed to learn to use a system, general costs and support needed have to be balanced against the added value of a new system. Cost of a system per patient case is a very important factor.

5.3.5. Relating results to trends in medical imaging and treatment

Acknowledging trends in the context of use can help identify requirements for a prospect system that aims to be applied in that context. For the VRE in particular trends in medical imaging and vascular treatment present useful information to requirements and constraints its clinical context of use poses.

The findings of this study generally fit in with the three trends in medical imaging procedures as reported in Sakas, (2002):

- Additional costs of better diagnosis alone are no longer accepted: most clinical pictures can be diagnosed sufficiently. Extra expenses are only accepted if the additional information has an impact on the therapy procedure. Related in the study is the efficiency requested and the expressed consideration that when extra tests and imaging are only applied if they are perceived as truly valuable to diagnosis and treatment success.
- Imaging procedures are increasingly used during interventions and not only during diagnosis. In this study especially possibilities for improved imaging during endovascular interventions were named.
- Moving from 2D to 3D: 3D information is gathered with new scanning techniques and this information is used. However, from this study it seems that predominantly still 2D representations are used. The findings suggest that, improvements in efficiency and accuracy of use of 3D reconstructions are needed.

Trends in vascular interventional treatments can be identified as well. Endovascular treatments are gaining in interest, more so than open bypass surgery, and are where suitable preferred; risk to patients is decreased and recovery is faster. Possibly due to this trend, but also due to the current practices of vascular treatment planning processes now, interest in the simulation of bypasses and blood flow showed by participants remained below expectation. The focus of utility of the VRE to aortic blood flow and bypass simulation seems less than ideal. Chances for supporting and aiding visualization during endovascular interventions are for example more apparent.

As explicitly stated by one participant, it would be very wise to involve the prospect users of the VRE more intensively in the further development process of the VRE to learn about existing needs and chances if the system is to actually be an asset to current work practices and fit its prospect context of use. The following chapter will include a discussion on the implications of findings from both this context analysis and earlier heuristic usability evaluation might have on the development process of the VRE as well as VR system development in general.

6. Discussion

This chapter will start with a discussion of the results of the VRE case study and the implications for further development of the VRE. The results of choosing heuristic usability evaluation and an ethnographic approach to context analysis as methodologies for this study are discussed. Subsequently, this chapter will elaborate on the insights that can be gained from this study for VR development in general. Application of context analysis in development of a VR system is elaborated upon as well.

6.1. Discussion of VRE case study results

This section will first highlight the main results of the VRE case study and discuss their implications on development of the VRE. Italic sentences refer to the research questions of this case study. It needs to be kept in mind that the goal of this study however is to add to VR research in general and after this section the general research questions of this thesis that pertain to VR development in general will be discussed.

6.1.1. Practical implications for VRE system development.

This section will discuss the implications of the case study for development of the VRE.
First, the research questions of the case study will be answered in order to in the end provide an answer to the questions whether *a Virtual Reality system such as the VRE would be suitable for supporting diagnosis and treatment planning in the proposed clinical context of use.* A selection of the case study questions and findings are concisely discussed below.

Does the functionality of the VRE adequately fulfil a user need or wish?
Additional systems will only be used if they provide an actual and perceived added value over existing systems and provide efficient and effective interaction. The current utility offered by the VRE does not quite fit current user needs:

Stereoscopic 3D visualization of vascular structures
While 3D information is used in scan reconstructions and 3D renderings are available, reconstructions are not usually viewed in 3D. Just offering 3D images is not likely to be very valuable to prospect users. Improved automatic 3D rendering possibly could be useful if they could provide efficient ways of working, with less manual input to construct a reliable 3D image. Chances do exist for reliable automatic measurement systems. Support for viewing multiple images and combining 2D and 3D views should be included. This combination has however been identified as a challenge when building a completely 3D oriented immersive application. Issues in 3D visualization and manipulation, such as difficulties in navigation and orientation, need to be dealt with.

Stereoscopic vision might provide greater insight in 3D structures, but in cases where the object's structure – in this case the abdominal aortic vessels- is relatively standard and has been encountered many times before by prospect users, actual value is unclear. It remains questionable if in such situations possible advantages actually outweigh the efforts of using equipment making stereoscopic vision possible. Existing literature cannot provide an answer for this particular situation.

Simulation of blood flow
Data on blood flow is used less intensively than expected. Simulation and measurements should be adapted to the user's needs and requirements. Functionality offering information that is perceived as unnecessary or already readily available is not appreciated. One of the main success-factors for a system like the VRE is the reliability of simulation results. Specifically in a medical context, simulation has to have been verified with real data. Directly offering simulation of blood flow within a vessel is not as interesting to participants as expected.

Simulation of the results of bypass placement

The simulation of immediate blood flow results after a bypass intervention as offered now are not interesting to prospect users. First of long term survival of a prosthesis is more interesting than the easier to predict immediate blood flow after an intervention. Additionally, while bypass surgery is still common place, the current focus of interest of interventional radiologists and vascular surgeons is not on bypass surgery, but more on recent developments in endovascular treatments. Support for these non-open interventions is more desirable.

Which usability issues challenge developing the VRE as a usable VR system?

- *What are the usability problems of the VRE prototype independent of a proposed context of use?*
- *What usability problems could potentially arise when the VRE would be used in a clinical context?*
- *Are solutions possible for these usability problems?*

A number of usability problems of the VRE prototype both in the laboratory and a clinical context have been identified. A lot of these problems could currently be encountered during development of any VR system. Satisfactory solutions are certainly not yet available for all of these problems. More extensive descriptions of the shortly listed problems below, have been described in the results section of this thesis and are not repeated in this section. Links to VR development and research in general are made.

Problems were apparent in the interaction modalities used in the VRE, especially in the magnetic tracking and voice recognition systems used. The prototype state of these elements of the VRE and technological imperfections caused usability problems. However, more principal than these issues, it is also unclear what kind of interaction modality, or combinations, are suitable for a certain type of interaction. Some researchers are providing the starting points for guidelines in this respect. Stedmon et al. (2004) for example state that speech recognition could be suited to discrete operations ('pick up') and not suited for tasks such as navigation. Another study (De Boeck, 2004) explored various options in combining interaction modes and has compared e.g. error rate and preference (seemingly not necessarily negatively correlated) in using haptics and speech recognition. Definite guidelines covering a wide range of input modalities and devices are not available however. This study cannot yet offer answers without dedicated experiments either and further research is needed.

The VRE suffered from ergonomic problems as well. Simulator sickness poses serious discomfort to some users. Ergonomics of input devices is still lacking as well. While attention to ergonomics of input devices is growing, little research is available of how to counter simulator sickness. Problems of simulator sickness might be difficult to counter (or perhaps in some cases even completely unsolvable). It is argued here however that such physical issues should be regarded as very serious, since they result in serious discomfort and can be very detrimental to the user experience of a system. These issues have to addressed, even though standard solutions are not (yet) available.

Some of these context independent usability problems are related to usability problems that could arise when applying the VRE in a clinical context. Simulator sickness for example is unacceptable for a surgeon who has to be comfortable and fully capable to perform interventions. Applying voice recognition and magnetic tracking will prove a problem as well. Ambient noise, beepers, frequent and valued cooperation and subsequent vocal discussions make using voice recognition hard. Metal objects and mobile clinical equipment make calibration of the magnetic tracking system difficult as well. Resulting inaccuracies could have very serious results if a system would be used for medical measurements.

What are the contextual utility and usability requirements and constraints for visualization and simulation systems aiming to support diagnosis and treatment planning for vascular disorders in a clinical context?

How diagnosis and treatment planning for vascular disorders are currently conducted has been extensively discussed in the results section of this thesis and will not be repeated here. For use of a system such as the VRE to be feasible in a clinical context of use, adjustments are necessary. Specialist systems that are already available for vascular treatment planning, seem to provide support where needed to these processes more effectively than the VRE now could. The VRE has to be adjusted to prospect users and their goals and tasks, as well as the technical, social, organizational and physical environment of users that could potentially affect system use. Utility of the VRE needs to be extensively adjusted and efficiency needs to be worked on before the system would be used in a clinical context.

A number of additional requirements and constraints have stood out in the study results. Certainly, more constraints can be identified, but listing these goes beyond the purposes of this thesis.

- Users
  Especially radiologists and surgeons have very little time to spare. Large numbers of patients have to be treated and little time is available for using planning system, nor is there for training to use such systems. Efficiency is extremely important. Long computation times and inefficient interaction via the user interface with systems are unacceptable. Their goal is to treat patient effectively, but also be efficient and economical in their treatment.
- Social and organizational environment
  It is vital that a system supports collaboration and permits vocal discussion. Responsibilities and hierarchies within prospective user groups should be respected. The system will have to conform to rules and standard governing medical practices. The added value of a system has to be confirmed by research and accepted by prospective users. Budget is constrained.
- Technological environment
  For a system to be successful it needs to be compatible with work practices and systems already in use, including printers. A need exists for incorporation of other electronic data and hardcopy information in a system. There is limited data storage available in hospitals and complete scan data sets may not be available.
- Physical environment
  Limited space is available for pieces of new large equipment. It has to be possible to use systems in hectic, dimmed conditions with background noises.

To summarize: effectiveness, efficiency and cost limitations are of great concern in the hectic environment of radiology and vascular departments and should be supported in the VRE's development.

Is a choice for a Virtual Reality system appropriate for processes in this context?
The usability problems of the VRE associated with choosing a virtual world-like implementation such as 'getting lost' in an environment that does not actually offer any information on the objects of interest, but also difficulty in combination of 2D and 3D images indicate that the currently used interaction techniques as well as the immersive VR paradigm of a virtual world might not be the best choice for the VRE applied in a clinical context. Especially since the value of using VR and stereopsis for the VRE application has currently yet to be proven.

These constraints are not only relevant to the VRE, but to more systems applied in similar contexts and those outside a medical environment. Many issues identified in this contextual analysis are similar to general considerations for system design offered in literature, such as Dix et al., (1998) and Preece et al., (2002). The identified issues are examples of information relevant to system development that can be gained from contextual analysis.

6.2. Discussion of methodology

The usability evaluation and context analysis for the VRE have resulted in a number of useful findings. However, these methods can pose challenges and have their limitations. Issues relating to the chosen methodology are elaborated upon here. First usability evaluation and context analysis in development of the VRE and other VR applications will be discussed. The effects of choosing this type of case study to gain more insight into the general research questions of this study will be discussed afterwards.

6.2.1. Discussion of usability evaluation of the VRE

When analysing the various VR evaluation methods, it was noticed that most of these methods focus on applications offering a ‘virtual world’. Examples of these are architectural walkthroughs, which aim to immerse the user in a complete environment. This fully immersive focus seems to make some of the heuristics in the literature (e.g. presence in the virtual world) less relevant to applications such as the VRE, in which user focus is on 3D objects. This might offer a first indication that the chosen interaction paradigm might not be ideal. However, no methods are available to determine if the VR paradigm is a right choice for an application and this question is not addressed in any heuristics. An exploration of what choices have to made in the future is deemed more useful here than a listing of all of the VRE’s usability problems.

While heuristics can serve as a useful reminder of what problems to look for in an application, in this evaluation a walkthrough of the application resulted in the identification of more problems than when focusing on using available heuristics. But even this walkthrough requires ‘stepping back’ and thinking about major design decisions that might play a major part in the occurrence of these problems. It has been realized that in this case, it would not be very productive to concentrate on finding every usability problem in this version of the VRE, when higher-level choices such as the choice for a certain type of interface (or multiple options) and what main functionality to include have not been satisfactory made. Whether 3D visualization and stereoscopic vision is beneficial for applications as the VRE remains to be studied as well.

Evaluation and context of use

Usability evaluation without a regard for the context of use is not likely to result in an overview of all usability problems that are likely to arise when the system will be implemented. A first indication of usability problems can be given, but many issues cannot be completely understood without exploring if a system would fit in a certain environment. In the VRE case study, this has been illustrated. For example, voice recognition problems that would arise when beepers that can go off at any time would not be taken into account, and by problems with using the VRE magnetic tracking system that currently needs to be calibrated to a particular environment; making mobile use of the system or equipment around the system difficult.

Not only such technological issues, but also questions of a more cultural nature have arisen. For example, the VRE offers a number of often-used synonyms for delete, including ‘kill’. While it may be useful that these synonyms are offered to users who are familiar with this term, using “kill” as a synonym for “delete” might not be appropriate to all sensitivities in a medical environment. While it might not be the most original example, and possibly might seem a overly squeamish one, it does illustrate that what is ‘acceptable’ might differ from developers to end-users and their context (although it needs to be noted that no survey has taken place to confirm this particular example).

It has become clear that to the VRE development team more information is needed on the context of use of the VRE and user requirements for such systems. A need to look into issues surrounding utility has indeed in this case been identified during usability evaluation as well. Usability evaluations often also address utility issues, as noted as well by e.g. Hilbert and Redmiles (2000). The heuristic evaluation resulted not only in the identification of potential usability problems, but also in the insight that it yet was very unclear what functionality would actually help support medical practitioners. A look at the context of use to gain insight in both desired functionality and possible context-dependent usability problems of the VRE would be necessary before any final usability evaluation could ever take place.

6.2.2. Discussion of VRE context analysis methodology

The heuristic evaluation resulted in the identification of a number of areas that needed attention in this case of development of the VRE system. Heuristic evaluation alone however, was not deemed sufficient to assess the usability and acceptability of the system for end-users. This study shows that carefully studying usability and context of use can offer crucial information for development of VR systems. Even a small-scale study such as this one, can be extremely useful in identifying needs and potential user-problems with the VR system in development.

The chosen ethnographic approach to contextual analysis combining interviewing with observation, proved to be useful for exploring context of use. The combination has yielded a larger variety of data and combines the advantages of the two. Many of the observations could not have been understood without the information gathered earlier in the interviews. Frequent comments by interviewees were, that they could not properly explain certain things without actually demonstrating carrying out these tasks. Most of the surprising pieces of information were gathered through informal discussion during observations. Since these could not be anticipated, these could not likely have been gathered during a formal interview. Which tasks should be observed in the study could only be found out by using the interviews to get an overview of the processes surrounding diagnosis and treatment planning.

Gathering artefacts during observation proved to be very useful in this study as well. All information resulting from tasks were seen to be recorded in communication reports and patient files, to see this items of information structured on a piece of paper is easier than trying to list them by asking participants, who might forget information or simply do not have the time to elaborate on the question. Of course, the details of the artefact's actual use have to be verified with participants. Photographing and concrete examples help communication with the development team and serve as external memory tools to the researcher. For the gathering of artefacts however permission is needed from participants, especially in a medical environment with privacy policies in place.

Challenges in contextual analysis

Context analysis offered useful information to support development of the VRE. Some considerations could have been identified using other methodologies in human-computer interaction such as pluralistic walkthroughs, prototype evaluation and user involvement as described by e.g. Preece et al., (2002) and Dix et al. (1998). However, important insights gained in this study to guide system design would unlikely be discovered without contextual analysis of the environment of use. It could therefore be recommended to perform some kind of context analysis in development of systems for a context of use a development team does not know much about, even if only a limited study is possible.

A limited study, instead of an extensive research endeavour, might even be somewhat imposed. It may be difficult to recruit enough participants for a study. E.g. medical practitioners for example are very busy and have quite important responsibilities. Possible issues of limited depth and representativeness can come into play; results cannot always be easily generalized. Limitations in methodology and resulting qualitative data have to be acknowledged.

Contextual analysis and ethnographic observation do offer a variety of other challenges as well. Context analysis takes considerable time and effort. Certainly the interviewer and observer should be trained in ethnographic methods, interviewing, observation and analysis. Even when the observer is skilled practical challenges can be encountered. Observing while taking notes at the same time in a time-critical environment as a lone observer is challenging. However, in certain settings such as in the medical context of this study using other tools as video or tape recording meetings might be unacceptable to participants. In general, when using such extra tools it needs to be kept in mind that analysing and recording any extra data takes more time and effort.

Other practical issues can come into play as well. For example, jargon is often used in a medical context. Certainly studying jargon of the context to be studied is recommended, but when the researcher happens to be unfamiliar with certain terms during observation, researchers should not be embarrassed to ask for these explanations. Especially in a medical context, many practitioners do not mind. Many are used to having students look over their shoulder anyway. For younger researchers being treated by participants as one of their students instead of a researcher might be a problem. While the researcher is there to learn from observing and interacting with participants, the researcher especially wants to see representative work being done in context, with some explanation where needed, but does not need to be lectured at. The researcher's goals have to be met and sometimes a leading role has to be played, as Rouncefield et al. (1997) puts it: ‘Respect the setting and its participants…Be courteous…but don’t be a doormat’.

Authors such as Robson (2002), Preece et al. (2002), Hix and Gabbard, (1998), Rouncefield et al., (1997) and Hughes et al., (1995) offer very useful advice specifically geared to ethnographic research and the sorts of problems that need to be tackled.

Incorporating context analysis in the development process

A defining success factor for contextual studies in system development is making sure that findings do actually feed back into the development process. Both benefits and limitations of the analysis should be clear to all involved. Using context analysis and ethnographic research however may be a relatively new experience to a technical development team, but in the end they need to work with the results of the study. Cultural differences between a more social research oriented observer and technically oriented developers posed a challenge in this study. Developers might not have great faith or experience in working with ethnographic research and might even be somewhat opposed to qualitative research in general. Some getting used to each other’s interests may be required. Explaining to each other why certain issues are important in requirement engineering and development to both of the sides is very useful. The methodology has to be clearly explained. The researcher needs to be flexible and sensitive to the development team’s questions and concerns. If possible and desired, these can be iteratively incorporated in the study. Social skills are not only needed in carrying out research, but as well in finding out what a study should entail and the communication of its results. Great attention has to be put on communicating results of the study to the people who need to work with these results. One final presentation of a report is not ideal. Structuring results is vital. Providing narrative scenarios of encountered situations can provide useful examples, but important issues need to be highlighted in a more structured way.

When using contextual analysis in iterative design, general problems of an iterative process have to be dealt with: it is hard to overcome early decisions on prototypes, hard to not just see symptoms but find underlying cause of usability problems (Dix et al, 1998). Investment of development team in costs, efforts and potentially even some emotional binding to a prototype is hard to overcome. The research might be more focused at finding a use for a system available, than on actually finding out what users need. A HCI expert is usually not completely independent from a development team and might depend on their support and might be reluctant to completely denounce even the worst of their designs. What a 'successful study' in the opinion of the development team might be very different from the goals a human factors researcher might have in terms of interest. As Bossen (2002) points out for the specific case of Contextual Design, contextual approaches might be called user centred, but perhaps they actually focus more on the needs of the development team. There can be tension between providing usable results to a development team, who want to find out how they can adjust and market their system, and a researcher who want to find out what users in a particular environment would actually need, which could be something else than the kind of systems the development team is able to build. The development team will have to actually pay attention to results of the contextual study for it to be useful. Beyond such a study, it is vital not to treat one completed user study as fulfilling some user involvement obligation and completely rely on these results, but to actively involve users in the whole development process. It has to be clear that a context analysis study will not result in a right-and-ready set of requirements that can be directly translated into a complete system design.

When to stop?
In the case of the VRE, extensive user involvement is crucial, and certainly more user research is needed than this study alone. However, even in this relatively small study, later observations mainly seemed to confirm earlier findings and little additional information could be gained. The value of additional ethnographic study needs to be balanced against the efforts, time and money. This poses a problem; additional value gained from possible future surprise findings cannot be predicted beforehand. In the clinical context that has been explored in this study, many work practices are bound by rules and standards pertaining to all Dutch hospitals. Even though differences between hospitals and departments exist and the context is complex, generalizations of findings is somewhat easier than in less formalized settings. In such rule-bound cases, after an exploratory context analysis, a shift could possibly be made to more participatory design situation with, quite conceivably less, participants who are prepared to invest considerable time and effort. In this way, more information that is actually applicable in design and implementation of a system can be gained more efficiently. Such choices however, have to be made for each project and development stage individually.

Adjusting contextual methods to the job at hand
As Bossen (2002) points out: there is probably no recipe that will fit all projects and contexts. Focus, plans and methods might have to be adapted to the situation at hand, and repeatedly adapted during a study as well. Strictly adhering to 'prescribed' methods should not be a goal in a study, getting the most valid and informative results from a study should be.

The approach to observation taken in this study of observation sessions in which clarifying questions are asked, resembles the Contextual Inquiry interview structure. Still, differences with the Contextual Inquiry approach are apparent. Since the Inquiry interview model of Contextual Design with questions asked during task performance itself, can be hard to apply in a medical setting. Some tasks are critical and cannot be interrupted. Asking questions at unobtrusive moments seemed to work fairly well and seemed to be less inconvenient to participants than a complete interview session. Some participants took it unto themselves to actively talk aloud during individual tasks, this is very useful, but simply cannot be expected from all participants. In those situations where talking aloud is possible, it should however be encouragement by the observer.

Many participants indicated they did not have the time for such a session and even when they did agree to an interview session, sessions were often interrupted or cut short. A delicate balance needs to be found between getting all the information a study would benefit from and the willingness of the participants to taking the time to answer questions, or to look up information and providing this information at all. Each new question takes time away from the participant's tasks, participants have their own limits, which should be respected. Aggravating participants might not be beneficial to their participation and willingness to share information, or provide contact information of other possible participants.

Such limitations cannot always be avoided and practical adjustments to methods can be needed and should not be avoided simply because they might not match existing 'ideal' methods.

Ethical and emotional considerations in a medical setting

Researchers that have no previous experience in a medical context should be aware of specific sensitivities of the field. The ethical and possible emotional implications of observation in medical settings with patients present or even getting to see personal information without seeing the actual patient in the flesh cannot and should not be ignored. While it may be convenient for the researcher to dismiss these issues in gathering as much information as possible and while this might be beneficial to development of useful tools, this might not be ethically just or even possible for the researcher. First of all, patient data should be kept private and this poses restrictions on for example collecting artefacts and photographing. In certain situations it could be argued that patients have to consent to being observed as well.

In this study this issue was kept implicit and participants asked consent to their patients, who in this study never objected to being observed. However, they often seemed to be under the impression that the observer was a medical student taking copious notes. This is not surprising considering the circumstances. Due to the somewhat dominant role of the medical practitioner and constant time pressure, there seemed to be little room for ample explanation and for getting truly informed consent. This is most pressing in cases of observing operations or in attending patient consultations. Medical practitioners do not and cannot take the time to explain why an observer is present. In this study, participants often introduced the researcher as a student without naming the field of study, or only with descriptions as 'someone who is looking over my shoulder today'. The practitioner takes the lead in meetings and spontaneously offering a lengthy explanation is inappropriate in most situations.

Additionally, when coming into contact with patients, the observer was required to wear a doctor's coat. Wearing this provokes assumptions and expectations of patients. It needs to be kept in mind that the patient might see the researcher as a representative of the hospital and the researcher cannot simply ignore remarks and questions. Participants, patients and their accompanying family or friends may actively address the observer. Questions and remarks may come as a surprise, especially when the researcher has chosen a fly-on-the-wall approach. Researchers do need to be honest when they are not medically trained and that they cannot offer any advice. Having a very short explanation for patients of who the researcher is and what the researcher is doing in the hospital should be ready in mind; time is precious, but clarity is important.

Emotionally, a certain discomfort can be experienced when a patient thanks the researcher at the end of a meeting, while the researcher feels that s/he has principally not offered anything to a distressed patient and should be thankful for being able to observe the consultation and being present at a potential very personal and emotional time for the patient. Patients getting bad news from the medical practitioner might have strong emotional reactions that can be challenging to observe in a neutral fashion. It might be difficult to the observer to remain neutral and not show feelings in emotionally charged situations or offer some sort of support to a patient, especially when the approach of the medical practitioner may seem sub-optimal, for whatever reason. Respecting the practitioners, the patient and others present, is vital.

The issue of possible confrontation with physical patient trauma comes into play as well in a medical setting. For example, anguish over wounds or worrying about doing something harmful by being in the operating theatre might blur the observer's focus. Observing an operation might be challenging and frightening for some, this can be reckoned with in advance by not putting that particular researcher in that situation. However, surprise confrontations with physical and emotional trauma are likely when observing a great number of surgical patient consults. To the author of this thesis, a particularly difficult situation entailed walking in on a consult of a surgeon with a very emotional patient, coupled with a totally unprepared sight of the patient's open amputation wound. Potentially disturbing a participant or patient with the researcher own emotional reactions is not beneficial and personal reactions should be suppressed. Some debriefing can be needed for the researcher, something that can to be dealt with and thought about before starting out.

Some inconveniences to participants, patients and researchers themselves can be avoided by extensive preparation. Simply measures, like for example not wearing extensive jewellery to an observation of an operation can save time and annoyance to participants. This author jumped at the surprise opportunity of seeing a bypass operation, which resulted in a vascular surgeon helping to take off tens of bracelets of the embarrassed author's wrists before the required pre-operational scrubbing could commence. Other issues could be thought of as well; suppose for example that the researcher slips and falls onto the patient on the operating table and extensive trauma results; who is accountable? If possible, thinking of situations the researcher might encounter and asking participants beforehand what reactions would be appropriate during e.g. patient contact, might help the researcher in avoiding uncomfortable situations, stress and potentially serious problems. Certainly, instructions of the medical staff should be adhered to.

6.2.3. The VRE case study methodology and general results

Some of the findings of the VRE case study might not be relevant to other applications, but some general issues can be identified. The VRE case study did result in insights about the usability challenges that are encountered when developing a system. Findings are more or less anecdotal, especially in regards to the problems VR system developers can encounter. These usability problems that could be solved using guidance from earlier literature are however of importance in general. These issues might indicate a certain immaturity of usability of VR systems and a lack of attention to human-computer interaction in VR research. Needs for further research can certainly be identified using such case studies, when examining what open questions cannot be answered using existing literature.

Context analysis can help to decide whether VR and which type of VR would be most applicable for a given situation, by providing more information about that context. Indications that certain more general guidelines are needed can be gained from case studies such as this one. However, evaluating a single system for a single specific context cannot provide complete guidelines matching properties of various types of contexts with e.g. suitable grades of immersiveness. Such guidelines on for example the utility of stereoscopic vision in particular situations have to be experimentally constructed. Multiple cases should be studied to gain a more complete overview of issues at play in VR development.

In the results of this study the need for guidance on finding out whether VR is applicable in a particular situation stands out. Also, guidelines for determining what type of VR is most suitable are needed. These needs are discussed in the following section.

6.2.4. When is choosing VR right and how to choose 'the right VR'?

The findings indicate that a desktop application would be more suitable in the explored context of use than a large completely immersive virtual environment. The reduction in hardware costs coupled with the greater familiarity of users with desktop computing make desktop VR a viable alternative to an immersive application. In the case of the VRE, it was concluded that a desktop application might be more suitable in a clinical setting. Findings suggest that in the domain of vascular diagnosis and intervention planning fully immersive systems offering simulated operating theatres might be more suited to teaching situations. Effectiveness of these systems for training purposes needs further research.

The findings offer an indication of factors that influence the effectiveness of a VR system and appropriate degree of immersiveness. Findings in literature have been quite limited. Mulder and Van Liere (2000) for example discuss a number of advantages of fish tank and desktop VR systems. These include simplicity to set up compared to large surround-screen projection based systems. Systems cost less, require less specialized components and are easier to transport. As of yet, fish tank VR systems can also provide higher resolutions and higher light intensity. Smaller fish tank VR systems can be used in more versatile ways, functioning as ordinary desk top computer systems, larger ones as display media for presentations. This is very useful in the VRE's prospect, where e.g. 2D and 3D images have to be alternated and meetings require presentations of text and 2D images as well. A major disadvantage of fish tank VR systems is their limited workspace, making them less suitable for walkthrough type applications. Fish tank VR can also in some cases be limited in depth perception of objects. Demiralp et al., (2003) have performed an anecdotal, very small-scale study comparing VR modalities resulting in similar and additional criteria. Their hypothesis is that for exocentric applications that focus on simulating an object, desktop VR (in our definitions) would be preferable to a CAVE.

Table 3 offers an overview of a number of other possible criteria that could be included in choosing a particular type of system. These are suggestions and examples of criteria as derived from the study results, this list is certainly not argued to be a complete overview.

	Desktop VR	*'Fish tank' VR*	*Fully Immersive VR*
Budget	Constrained budget.	Relatively constrained budget.	Ample budget.
Focus	Exocentric focus (on objects)	Exocentric focus (on objects).	Focus on exploring environments.
Space	No room for specialized equipment / transport desirable	Physical space available for a limited volume of extra equipment / transport desirable.	Ample dedicated physical space available.
Time	No extra time can be spend on using specialized equipment and complex manipulation. Hectic context of use, users need to switch tasks quickly.	Insight into a complex structure is needed and decision speed is less vital.	Insight into a complex structure is needed and decision speed is less vital, dedicated context of use can be supplied.
Training	Users are less inclined to learn using new equipment and technology.	Users are willing to learn using new equipment and technology.	Users are willing to learn using new equipment and technology.
Compatibility and integration	Compatibility and integration of different types of data is very important and not attainable with current immersive environments.	Compatibility and integration of different types of data is sufficiently attainable with fish tank/ immersive environments.	Compatibility and integration of different types of data is sufficiently attainable with immersive environments.
Stereoscopy	Stereoscopic visualization and immersion do not significantly add to task performance, or 3D images would have to be combined with 2D images.	Stereoscopic visualization and immersion significantly add to task performance.	Stereoscopic visualization and immersion significantly add to task performance.
Use in education	For individual 'tutorial' training, for example at home (Zudilova and Sloot, 2003).	For individual training when e.g. haptic feedback is of value.	Education in lecture modes (Zudilova and Sloot, 2003).
Collaboration	Quick, informal collaboration of a small user group needs to be supported.	Collaboration of large groups is not necessary. Very small groups can be supported.	Collaboration can be supported by ample interaction devices and physical room.

Table 3. Possible criteria for choice of VR environment and degree of immersion..

Future research is hoped to provide more insight into these and new criteria so that a concise set of guidelines can be developed.

Parallels can be drawn to other guidelines offered in literature on human-computer interaction. General (ergonomic) guidelines on system development are offered by for example Dix et al., (1998). Guidelines on the design of virtual environments have been offered as VR-specific usability heuristics by e.g. Kalawsky (1999), Steed and Tromp (1998), Kaur et al., (1999), Sutcliffe and Kaur (2000), Neale and Nichols (2001) and Willans et al., (2001). These usually do not deal with choosing a particular kind of VR. When research on stereoscopic vision is evaluated, it needs to be kept in mind that even if stereoscopy is proven to be useful for gaining insight in particular kinds of 3D objects and environment, it might not be useful for other situations. To draw on medical examples; stereoscopic vision of intricate vascular structures in the brain, which are not well-known to physicians, stereoscopic vision might be more beneficiary than in assessing the condition of the abdominal aorta, the general structure of which is well known. Beyond such limits of scope, it also needs to be kept in mind that guidelines have other limitations, (see e.g. Dix et al., 1998 or Van Welie, 1999); new technologies can also render specific guidelines obsolete, or new ones necessary, in the rapid developing field of VR this should be considered.

6.2.5. Focusing VR research

There does not seem to be a clear focus of where the VRE system will be heading and various scenarios are often talked about; VRE training systems, 3D scan visualization, diagnosis and planning systems. This, while an immersive surgical simulator used as a training system to simulate complete operations has other requirements than a planning system or a scan visualization tool (which have been outlined in chapter 4). Now that a clinical context of use has been studied in an exploratory way, it is more feasible to make a more informed choice on what type of system is to be developed. In participation with prospect users, choices can also be made about focusing on for example building a system for planning endovascular interventions instead of bypass operations. Since the idea of an immersive VR system in a clinical context might not be realistic, instead surgical simulators could be built for training purposes, instead of developing such systems for a clinical context.

However, in case of the VRE development team, a decision to focus research instead of diverging into multiple directions, could be useful to get results from the VRE research in a faster and more economical way. Since development for end-users requires a lot of expertise on human-computer interaction, it seems likely that a more technological team will decide to focus on delivering technology and simulation algorithms and abandon the idea of actually making systems for end-users altogether. Investments can then be made to further develop the chosen own area of expertise and provide technology to organisations which aim to make systems for end-users.

Convergence in general VR research

Developing a VR system requires knowledge and skills on a wide variety of fields. For example, technological implementation, human-computer interaction issues and knowledge on the context of use, users and their tasks are needed. Each of these areas in turn are complex as well. It seems that some VR development teams insist on building complex VR systems completely on their own, while convergence could possibly be more beneficial to development of their systems and more valuable to VR research in general as well. Questions of either leaving particular issues up to other specialist teams, or trying to develop an extra aspect of the system in an own department have arisen during development of the VRE as well. For example, trying to apply haptics in the VRE system is thought about for possible surgical simulation applications. However, there is a trade-off between having the option of investing own manpower in haptic research. This, while other research groups have more expertise in this field.

Considerable efforts are required to acquire the knowledge and skills on a wide variety of fields, while the results even might not be optimal since becoming or including experts on ‘everything’ in VR development is almost impossible. Getting advice or possibly buying components (or expertise) from others, while focusing own efforts on the ‘core business’ of a research group could be a more effective and efficient option. Insisting on building a system on their own and starting experimenting with new technologies from scratch can increase time and costs needed to implement a system. It is argued here that various specialist groups should take advantage of the possibilities of sharing and benefiting from each other's expertise and should not try to 're-invent the wheel'. A similar point is made by e.g. Montgomery et al., (2001) as well. More efforts can thus be invested to develop own areas of expertise and trying to answer the open questions in those areas, while only extending to other fields when actually useful and feasible. Convergence should not mean trying to get expertise on every aspect individually by including any expert in a research group, it means working together towards a common goal, e.g. benefiting from each other’s expertise for individual systems but also building up the general knowledge base of the VR research community as a whole.

7. Conclusion

As indicated by the findings of this study, the VRE as yet cannot be applied in a clinical context. The system has not been developed in accordance with needs and requirements of its prospective user groups, tasks and context. A number of areas have been identified where the VRE did not match user requirements and could not be implemented as is in a clinical context of use. The usability evaluation and context analysis proved to be very useful in this respect.

However, the goal of this thesis was not just to add to development of the VRE, more general goals were aimed for. The following general research questions have been asked in the problem statement of this thesis:

- What utility and usability issues exist for utilization of VR systems in a real world context?
    - Which usability issues make developing usable VR systems challenging?
    - Do VR systems provide sufficient added value to warrant their use in spite of potential drawbacks?

Quite a number of usability issues have been identified during this study. Psychological factors, immersion and stereoscopy, navigation and orientation, physical effects and ergonomics, (multimodal) interaction in VR systems; all provide chances for improvement. Available usability evaluation methods do not all accommodate all of these issues, but they do offer useful guidance in finding such usability problems independent of the prospect context of use.

The usability and utility problems that are caused by insufficient attention to users and their context of use depend on the particular context at hand. A VR system that offers utility that does not fit user needs will probably not be used. More extensive user involvement and applying contextual methods as in this thesis to find out more about their context of use could produce more usable VR systems that offer functionality that is useful to the prospect users.

More general guidelines are however needed to fit VR systems and interaction devices to users, their tasks and environment. Many of the identified usability problems are (partially) caused by the immaturity of VR technology, or supporting technology used in these systems, as illustrated by e.g. voice recognition failures in the VRE. Improvements can be expected in these areas. More principally however, problems can also be caused by using a virtual environment for a particular application. Offering a virtual world when unnecessary for example, requires navigation making the application inefficient. Potential drawbacks have to be balanced by the value of using of a VR system in a particular situation.

Whether VR systems provide sufficient added value to warrant their use in spite of potential drawbacks is a rather daunting question to answer. Various VR systems have proven their use on multiple occasions; chapter 2 provides examples in a medical environment of surgical simulators used in education of students, or use in distraction-by-entertainment for burn victims while dressing their wounds. In the VRE case though, the question has to be negatively answered. The problems found in the usability evaluation of the VRE and the potential problems in a context of use, coupled with the changes that would be needed to the system's functionality and the uncertainty whether a VR system would be applicable at all, make application of the VRE in a clinical context very questionable. Ultimately, the fit of a VR system depends on the particular user needs and properties of the particular context of use. It has to be examined whether VR is useful for supporting the specific user tasks in that environment. A more general statement can however be made.

Currently, there can be a tension between making a usable system and pursuing research interests. The technology push and experimental approaches with which many VR systems are currently being developed, have their merits in development of new technology. However, with such a focus on technological feats with little involvement of prospect end-users and limited attention to context of use, usability and utility problems are to be expected. Certainly, conflicting goals exists between a researcher pursuing studies into a particular system paradigm, and real world demands of providing usable systems. The exclamation "but if we change that, it wouldn't be a real VR system any more" from a developer during a casual conversation about restricting navigation in the VRE to prevent getting lost in the virtual environment, illustrates this conflict.

It could be argued that, the time has come to match VR to actual types of applications where the paradigm is useful and not just 'try out' the paradigm for individual applications regardless of whether the paradigm is actually useful in that situation. Then VR systems can be matched to those types of situations where they actually provide sufficient value to warrant their use.

7.1. Future research

Future developments of the VRE will include more attention to usability issues and experimentation with various types of VR systems. Currently desktop and 'fish tank' VRE versions are in development to try and cater for user needs and requirements in varying contexts. Future research should investigate which evaluation methods are most useful for application to VR systems and which methods are most applicable in exploring context of use. Research is needed to gain more insight into real-life needs and requirements for systems such as the VRE. Future research will take into account the limitations of this study. Development of the VRE will include more representative user involvement (Zudilova and Sloot, 2003). More detailed contextual analysis and quantitative analysis is planned into the general issues this study raises. The research described here has been carried out for a specialised application area, but the methodology and its results can be applied more broadly.

7.1.1. Research possibilities for VR development

A number of (unresolved) issues make developing usable VR systems challenging. Usability issues involving 3D interaction, such as manipulation of 3D objects and navigation problems in virtual world deserve attention. Immaturity of interaction technology and input devices make building a usable VR system challenging as well. The choice between various input modalities, voice recognition and devising possible usable combinations between those modalities is daunting.

However, before trying to better a VR system in this respect, it has to be made sure the VR paradigm is actually applicable. The discussion of this thesis provides some starting points for this question and the subsequent choice between types of VR systems, but certainly more research is needed. This includes research on the uses of stereoscopy for particular tasks in various contexts.

Such future research can help proliferation of more usable VR systems to situations in which they are actually useful and fulfil a user need. This way, the perception and utility of VR systems can change from being technological experiments into systems that are actually valuable to a wider audience. Constructing a usable VR system is by no means yet a straightforward process; many open questions in VR research remain. However, more extensive attention to usability and context of use could provide for more usable systems that are actually applicable in 'the outside world'.

References

Abran, A., Khelifi, A., Witold S., Seffah, A. (2003) Usability Meanings and Interpretations in ISO Standards, Journal of Software Quality, 11 (4), 325-338.

Balaguer, F., Mangili, A. (1991) Virtual Environments, *New Trends in Animation and Vizualization,* Wiley professional computing, Chistester, 91-105.

Bajura, M., Fuchs, H., Ohbuchi, R. (1992) Merging virtual objects with the real world: Seeing ultrasound imagery within the patient, *Computer Graphics*, *26(2)*, 203-210.

Bell, G. (2001) Looking Across the Atlantic: Using Ethnographic Methods to Make Sense of Europe, *Intel Technology Journal,* Third Quarter 2001.

Belleman, R.G. (2003) *Interactive Exploration in virtual environments*, PhD thesis, University of Amsterdam, The Netherlands.

Belleman, R.G., Kaandorp, J.A., Dijkman, D., Sloot, P.M.A. (1999) GEOPROVE: Geometric Probes for Virtual Environments, Proceedings of the HPCN Europe 1999, Amsterdam, The Netherlands, 817-827.

Belleman R.G., Sloot, P.M.A. (2001) Simulated Vascular Reconstruction in a Virtual Operating Theatre. Proceedings of the CARS 2001 conference, Berlin, Germany.

Bevan, N. (1995) Measuring usability as quality of use, *Software Quality Journal*, 4, 115-150.

Bevan, N. (2001) International standards for HCI and usability, *International Journal of Human-Computer Studies*, 55, 533-552.

Bevan, N., Kriakowski, J. Maissel, J. (1991) What is usability?, *Proceedings of the 4th International Conference on HCI*, Stuttgart.

Bevan, N., Macleod, M. (1994) Usability measurement in context, *Behavior and Information Technology*, 13 (1 and 2), 132-145. http://www.usability.serco.com/papers/music94.pdf, last visited 10th December 2003.

Beyer, H., Holtzblatt, K. (1998) *Contextual Design: Defining Customer-Centered Systems*, Morgan Kaufmann, San Francisco, United States of America.

Blechner, M. , Monaco, V., Knox, I. (2003) Using Contextual Design to Identify Potential Innovations for Problem Based Learning, *Proceedings of the AMIA Symposium*. http://www.health.pitt.edu/users/rebecca/Publications/Blechner_2003_AMIA.pdf, last visited 17th November 2003.

Bossen, C. (2002) Ethnography in design: tool-kit or analytic science, Proceedings of PDC 2002, Malmö, 338-343.

Bowman, D.A., Gabbard, J.L., Hix, D. (2002) A survey of usability evaluation in virtual environments: classification and comparison of methods, *Presence*, 11(4), 404-424.

Bowman, D., Johnson, D., Hodges, L. (2001) Testbed Evaluation of Virtual Environment Interaction Techniques. *Presence: Teleoperators and Virtual Environments,* 10(1), 75-95.

Bowman, D., Hodges, L. (1999) Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments, *The Journal of Visual Languages and Computing*, 10(1), 37-53.

Bricken, W. (1990) Virtual Reality: Directions of Growth, *Notes from the SIGGRAPH '90 Panel*, International Conference on Computer Graphics and Interactive Techniques, Dallas, Texas, United States, August 06 - 10, 1990.

Brown, J., Montgomery, K., Latombe, J.C., Stephanides, M. (2001) A Microsurgery Simulation System, Proceedings of Medical Image Computing and Computer-Assisted Interventions (MICCAI 2001), Utrecht, The Netherlands, October 14-17, 2001.

Chesher, C. (1994) Colonizing Virtual Reality Construction of the Discourse of Virtual Reality, 1984-1992, *Cultronix,* 1(1).

Cobb, S.V.G., Nichols, S.C., Ramsey, A.R. and Wilson, J.R. (1999) Virtual Reality Induced Symptoms and Effects. *Presence: Teleoperators and Virtual Environments*, 8(2).

COVEN (1997) Usage evaluation of the initial applications, *COVEN deliverable 3.3*, University College London, United Kingdom.

Crabtree, A. Rodden, T, Mariani, J. (2002) Designing Virtual Environments to support cooperation in the real world, *Virtual Reality*, 2002(6), 63-74

Cuypers, Ph.W.M. (2001), *Endovascular treatment of abdominal aortic aneurysms*, PhD thesis, University of Utrecht, The Netherlands.

De Boeck, J., Raymaekers, C., Coninx, K. (2004) Improving haptic interaction in a virtual environment by exploiting proprioception, *Proceedings of the Workshop on Virtual Reality Design and Evaluation Workshop*, 22-23 January 2004, Nottingham, United Kingdom.

DFN Deutsches Forschungsnetz (1999) *DFN-Expo 1999 website*, http://www.dfn-expo.de/, last visited August 2004.

Difede, J., Hoffman, H.G. (2002) Virtual reality exposure therapy for World Trade Center post-traumatic stress disorder: a case report. *CyberPsychology and Behavior*, 5(6), 529-535.

Dinsmore, M. et al. (1997) Virtual reality training simulation for palpation of subsurface tumors, *Proceedings of the IEEE VRAIS '97*, 54-60.

Dix, A., Finlay, J., Abowd, G., Beale, R. (1998) *Human-Computer Interaction*, Second Edition, Prentice Hall Europe, United Kingdom.

El-Khalili, N. (1999) Surgical Training on the world wide web, PhD thesis, School of Computer Studies, University of Leeds, United Kingdom.

Emerson, T, Prothero, J., Weghorst, S. (1994) Medicine and Virtual Reality: A Guide to the Literature (MedVR) , HITL Technical Report No. B-94-1, *Artificial Intelligence in Medicine*, 6, 335-349.

Emmelkamp, P.M.G. et al. (2001) Virtual Reality Treatment in Acrophobia: A Comparison with Exposure in Vivo, Cyberpsychology and Behavior, Vol.4, No.3, pp.335-341, June 2001.

Forsythe, D.E. (1999) “It’s just a matter of common sense”: Ethnography as invisible work, *Computer Supported Cooperative Work*, 8, 127-145.

Fouskas, K.G., Pateli, A.G., Spinellis, D.D., Virola, H. (2002), Applying Contextual Inquiry for Capturing End-users Behaviour Requirements for Mobile Exhibition Services, *Proceedings of the 1st International Conference on Mobile Business*, Athens, 8-9 July 2002.

Gabbard, J. L. (1997) *A taxonomy of usability characteristics in virtual environments*. M.S. Thesis, Virginia Polytechnic Institute and State University.

Gabbard, J.L., Swartz, K., Richey, K., Hix, D. (1999) Usability Evaluation Techniques: A Novel Method for Assessing the Usability of an Immersive Medical Visualization VE, *Prcoeedings of the VWSIM '99*, San Fransisco, United States of America, 165-170.

General Electric (2004), *Advanced Vessel Analysis AW-4.0* http://egems.gemedicalsystems.com/geCommunity/aw/FlexTrial/aw4_0/eflextrial_pages/aw40_ava.jsp?pkg=ava&awVersion=AW40, last visited 29 august 2004.

Gobbetti, E., Scateni, R. (1999) "Virtual Reality: Past Present and Future," Sardinia, Italy: Center for Advanced Studies, Research and Development.

Grudin, J. (1992) Utility and usability: Research issues and development contexts , *Interacting with Computers*, 4(5).

Hackbarth, G., Grover, V., Yi,M.Y. (2003) Computer playfulness and anxiety: positive and negative mediators of the system experience effect on perceived ease of use, Information and Management, 40(3), 221-232.

Hackos, J., Redish, J. (1998) *User and Task Analysis for Interface Design*, John Wiley and Sons, Inc.

Haik, E., Barker, T., Sapsford, J., Trainis,S. (2002) Investigation into Effective Navigation in Desktop Virtual Interfaces, *Proceedings of the 7th International Conference on 3D Web Technology*, February 24- 28, 2002, Tempe Mission Palms, Tempe, Arizona, United States of America.

Hand, C. (1994) Other Faces of Virtual Reality. *East-West International Conference on Multimedia,Hypermedia and Virtual Reality*. Moscow, Russia. http://citeseer.nj.nec.com/hand94other.html

Hemmings, T., Crabtree, A. (2002) Ethnography for Design?, Proceedings of the International Workshop on "Interpretive" Approaches to Information Systems and Computing Research, London, United Kingdom, 122-124.

Hilbert, D.M., Redmiles D.F. (2000) Extracting Usability Information from User Interface Events, *ACM Computing Surveys* , 32 (4), 384-421.

Hinckley, K. (1996) *Haptic issues for virtual manipulation*, PhD thesis, School of Engineering and Applied Sciences, University of Virginia, United States of America.

Hinckley, K., Pausch, R., Goble, J.C. (1994) A Three-Dimensional User Interface for Neurosurgical Visualization, *Proceedings of the SPIE 1994 Conference on Medical Imaging*, 126-136.

Hix, D., Gabbard, J.L. (1998) Usability Engineering of Virtual Environments**,** Stanney, K.M. (Ed.) Chapter 39, *VE Handbook*, http://vehand.engr.ucf.edu/handbook/, last visited 28th October 2003.

Hodges, L.F., Rothbaum, B.O., Kooper, R., Opdyke, D., Meyer, T., North, M., de Graaff, J.J., Williford, J. (1995) Virtual environments for treating the fear of heights, *IEEE Computer*, 28(7), 27-34.

Hoffman, H.G., Patterson, D.R., Carrougher, G.J., Sharar, S. (2001) The effectiveness of Virtual Reality based pain control with multiple treatments. Clinical Journal of Pain, 17, 229-235.

Hughes, J., King, V., Rodden, T., Andersen, H. (1994) Moving out from the control room: Ethnography in system design, *Proceedings of the Conference on Computer Supported Cooperative Work* (CSCW'94), 429-439.

Hughes, J., O'Brien, J., Rodden, T., Rouncefield, M., Sommerville, I. (1995) Presenting ethnography in the requirements process, *Proceedings of second IEEE international symposium on requirements engineering 1995*, York, United Kingdom , 27-35.

Hullfish, K.C. (1996) Virtual Reality Monitoring: How Real Is Virtual Reality?, Master's Thesis, University of Washington, United States of America. http://www.hitl.washington.edu/publications/hullfish/home.html,last visited October 20th, 2003

International Standards Organization (1998) ISO 9241-11:Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.

Jolesz, F. et al.(1997) Image-guided Procedures and the Operating Room of the Future, *Radiology*, 204, 601-612.

JTAP (2000) JTAP Project 305 Report; Human Factors Aspects of Virtual Design Environments in Education Project, JISC Technology Applications Programme (JTAP), Loughborough University, Loughborough, United Kingdom.

Kaiser et al. (2003) *Nederlands Huisartsen Genootschap Standaard Perifeer arterieelvaatlijden* , http://www.artsennet.nl/nhg/standaarden/M13/std.htm, last visited April 29th, 2003

Kalawsky,R.S. (1999) VRUSE - a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems, *Applied ergonomics*, 30(1), 11-25

Kalawsky, R.S., Bee, S.T., Nee, S.P. (1999) Human Factors Evaluation Techniques to Aid Understanding of Virtual Interfaces, *BT Technology Journal*, 17, 128-141.

Karat, J. (1997) Evolving the scope of user-centered design, Communications of the ACM, 40(7), 33-38.

Kassell, N. J. D. III, Graves, B. (1997) Telepresence in Neurosurgery: The Integrated Remote Neurosurgical System, *Proceedings of Medicine Meets Virtual Reality 5*, January 1997.

Kaur, K. (1998) *Designing Virtual Environments for Usability*, Ph.D. Thesis, City University, London, United Kingdom.

Kaur K., Sutcliffe, A., Maiden, N. (1999) A design advice tool presenting usability guidance for virtual environments, *Proceedings of the Workshop on User Centred Design and Implementation of Virtual Environments*, University of York.

Klinger, E. et al. (2002) Issues in the Design of Virtual Environments for the Treatment of Social Phobia , Proceedings of the 1st International Workshop on Virtual Reality Rehabilitation, Lausanne, Switzerland, November 7-8, 2002, 261-273.

Ko, A.J. (2003) A Contextual Inquiry of Expert Programmers in an Event-Based Programming Environment, *Poster presentation at CHI 2003*, April 5-10, 2003, Ft. Lauderdale, Florida, United States of America.

Kotonya, G., Sommerville, I. (1998) R*equirements Engineering: Processes and Techniques*', John Wiley and Sons, Inc, New York, United States of America,
as quoted in Preece et al (2002).

Lindeman, R.W., Sibert, J.L., Hahn, J.K. (1999) Towards usable VR: An empirical study of user interfaces for immersive virtual environments, *Proceedings of CHI'99*, 64-71.

Löwgren, J. (1995) Perspectives on usability. Technical Report LiTH-IDA-R-95-23, IDA, Linköping University, Sweden, http://www.ida.liu.se/labs/aslab/groups/um/publications/R-95-23.PDF

Marsh T., Wright, P., Smith, S., Duke, D. (1998) A shared framework of Virtual Reality, *Proceedings of UK-VRSIG'98*, Exeter, United Kingdom.

McNeill, M.D.J., H. Sayers, S. Wilson, Mc Kevitt, P. (2002) A spoken dialogue system for navigation in non-immersive virtual environments, *Computer Graphics Forum*, 21(4), 713-723.

Milgram, P. , Kishino, F. (1994) A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, Vol E77-D, No.12, December 1994.

Montgomery, K., Stephanides, M., Schendel, S., Ross, M. (2001) User Interface Paradigms for VR-based Surgical Planning: Lessons Learned Over a Decade of Research, Technical Report, Biocomputation Center, Stanford University.

Mourouzis, A., Grammenos, D., Filou, M., Papadakos, P., Stephanidis, C. (2004) Case Study: Sequential Evaluation of the Virtual Prints Concept and Pilot Implementation, *Proceedings of the Workshop on Virtual Reality Design and Evaluation Workshop*, 22-23 January 2004, Nottingham, United Kingdom.

MR-TIP (2004) MR-Technology Information Project, http://www.mr-tip.com/, last updated 8th August 2004, last visited 9th August 2004.

Mulder, J.D.,Van Liere, R. (2002) Personal Space Station: Bringing interaction within reach, *Proceedings of the VRIC 2002*, Laval.

Mulder, J.D., Jansen, Van Rhijn, A. (2003) An Affordable Optical Head Tracking System for Desktop VR/AR Systems, *Proceedings of the Immersive Technology and Virtual Environments Workshop 2003*, Zurich, Switzerland.

Nagel, E., Van Rossum, A.C., Fleck, E. (2002) Kardiovaskuläre magnetresonanz-tomographie, Steinkopff-Verlag, Darmstadt, Germany.

NASA JPL (2001) URBIE Urban Robot Project,
http://robotics.jpl.nasa.gov/tasks/tmr/stairclimbing.html, last updated 31st August 2001.

NCSA National Center for Supercomputing Applications (1999) *The CAVE at NCSA*, http://cave.ncsa.uiuc.edu/, last visited May 2004.

Neale and Nichols, S. (2001) Theme-based content analysis: a flexible method for virtual environment evaluation, International Journal of human-computer studies, 55, 167-189.

Nederlandse Hartstichting (2004) Nederlandse Hartstichting website, http://www.hartstichting.nl/, last visited 2nd August 2004

Nielsen, J. (1993) *Usability Engineering*, Morgan Kaufmann Publishers Inc., San Francisco, United States of America.

Normark, M. (2002) Some thoughts on ethnography for design, *Position paper for ACM 2002 Conference on Computer Supported Cooperative Work*, November 16-20, 2002, New Orleans, Louisiana, United States of America.

Pasman, G. (2003) *Designing with precedents*, PhD-thesis, Technische Universiteit Delft, http://www.io.tudelft.nl/id-studiolab/dwp/intro.html, last visited 17th November 2003.

Pausch, R., Burnette, T. (1995) Navigation and Locomotion in Virtual Worlds via Flight into Hand-Held Miniatures, *SIGGRAPH '95 Proceedings of the 22nd annual conference on Computer graphics and interactive techniques***,** August 6-11, 1995, Los Angeles, United States of America.

Poston, T., Serra, L. (1996) Dextrous virtual work, *Communications of the ACM*, 29, 37-45.

Preece, J., Rogers, Y., Sharp, H. (2002) *Interaction design: Beyond human-computer interaction*, John Wiley and Sons, New York, United States of America.

Reachin Technologies (2004) *Reachin Technologies homepage*, http://www.reachin.se, last visited August 2004.

Robson, C. (2002) *Real World Research Second Edition*, Blackwell Publishers, Oxford, United Kingdom.

Rouncefield, M., Hughes, J.A., O'Brien, J. (1997) *Ethnography. Some practicalities of Ethnographic Analysis*, Department of Computing Cooperative Systems Engineering Group Technical Report CSEG/27/1997, University of Lancaster, Presentation and workshop, University of Karlsrona-Ronneby, Sweden.

Ruddle, R.A., Lessels, S. (2004) Three levels of metric for evaluating wayfinding, *Proceedings of the Virtual reality design and evaluation workshop*, Nottingham, United Kingdom.

Satava, R.M. and Jones, S.B. (1999) Chapter 55: Medical applications of Virtual Reality, *Handbook of Virtual Reality*, Lawrence Erlbaum Associates, Inc., http://vehand.engr.ucf.edu/handbook/, last visited October 20th 2003.

Schreiber, A. Th. , Akkermans, J. M. Anjewierden, A. A. De Hoog, R. , Shadbolt, N. R. Van de Velde, W. (1999) *Knowledge Engineering and Management: The CommonKADS Methodology*, Cambridge MA, MIT Press.

Serra, L., Waterworth, J.A. (1996) Designing Virtual Selectors for Surgeons, *Applied Ergonomics*, 28 (4), 269-275.

Sharp, H., Finkelstein, A., Galal, G. (1999) Stakeholder Identification in the Requirements Engineering Process, *Proceedings of the 10th International Conference and Workshop on Database and Expert Systems Applications (DEXA)*, Florence, Italy.

Slater, M. (1999) Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire, Presence, vol.8, no.5, 560-565.

Slater, M., Usoh, M. (1994) *Body Centred Interaction in Immersive Virtual Environments*, in: (Magnenat Thalmann and Thalmann, eds) Artificial Life and Virtual Reality, J.Wiley, Chichester, 1-10.

Sloot P.M.A., Van Albada G.D., Zudilova E.V., Heinzlreiter P., Kranzlmüller D., Rosmanith H., Volkert J. (2003) Grid-based Interactive Visualisation of Medical Images, *Proceedings of the 1st European HealthGrid Conference*, Lyon, France, 57-67.

Smith, S., Marsh, T., Duke, D., Wrigth, P. (1998) Drowning in immersion, *Proceedings of the UK VRSIG'98*. UK Virtual Reality Special Interest Group.

SRI (2004) *SRI International Medical Product Development*, http://www.esd.sri.com/med_devel/telepresence.html, last visited August 2004.

Stanney, K.M., Mourant, R.R., Kennedy, R.S. (1998) Human Factors Issues in Virtual Environments, *Presence*, 7(4), 327-351.

Stedmon, A.W., Griffiths, G., Bayon, V. (2004) Single or Multi-User, Manual or Speech Input? An Assessment of De-coupled In Virtual Environments, *Proceedings of the Workshop on Virtual Reality Design and Evaluation Workshop*, 22-23 January 2004, Nottingham, United Kingdom.

Steed, A. (1993) *A Survey of Virtual Reality Literature*, Department of Computer Science, Queen Mary and Westfield College, Tech. report 623.

Steed, A., Tromp, J.G. (1998) Experiences with the Evaluation of CVE Applications, *Proceedings of Collaborative Virtual Environments Conference (CVE'98)*, 17-19 June, Manchester.

Sutcliffe, A.G. and Kaur, K.D. (2000) Evaluating the usability of Virtual Reality user interfaces, *Behaviour and Information Technology*, 19(6), 415-426.

Sutherland, I.E. (1968) A head mounted three dimensional display, *Proceedings of the fall joint computer conference (AFIPS)*, 33(1), 757-764.

Swanborn, P.G. (1997) *Basisboek Sociaal Onderzoek*, Boom, Amsterdam, The Netherlands.

Travis, D. (2003) *The bluffer's guide to ISO 9241*, Userfocus ltd, United Kingdom. http://www.userfocus.co.uk/articles/ISO9241.pdf, last visited 7th December 2003

Van der Heyden, J.E., Inkpen, K.M., Atkins, M.S., Carpendale, M.S.T. (1999) A User Centered Task Analysis of Interface Requirements for MRI Viewing, *Proceedings of the 1999 conference on Graphics Interface*, 18-26.

Väänänen-Vainio-Mattila, K., Ruuska, S. (1998) User needs for mobile communication devices: requirements gathering and analysis through contextual inquiry, *Proceedings of the first workshop on Human Computer Interaction with Mobile Devices HCIMD'98*, Glasgow, Scotland.

Von Wiegand, T.E., Schloerb, D.W., Sachtler, W. L. (1999) Virtual Workbench: Near-Field Virtual Environment System with Applications , *Presence: Teleoperators and Virtual Environments*, 92-519.

Van Welie, M., van der Veer, G.C., Elins, A. (1999) Breaking Down Usability, Proceedings of INTERACT 1999, Edinburgh, August 1999, 613-620.

VR Solutions (2004) 5DT HMD 800, www.vrsolutions.com.au/ vrsprod5dthmd.htm, last visited 24th March 2004.

Ware, C., Arthur, K., Booth, K.S. (1993) Fish tank Virtual Reality. *InterCHI '93 Conference on Human Factors in Computing Systems*, Amsterdam, 37-42.

Waterworth, J.A. (1999) Virtual Reality in Medicine: A Survey of the State of the Art, Umeå University, Sweden, http://www.informatik.umu.se/~jwworth/medpage.html, last updated July 1999, last visited October 20th 2003

Wever, J.J. (1999) *CT angiography follow-up after endovascular aortic aneurysm repair*, PhD thesis, University of Utrecht, The Netherlands.

Wildermuth, S., Thonier, G., Bruyns, C., Montgomery, K.. (2001) An Augmented Reality System for Endoluminal Aortic Stent Placement, Proceedings of the Medicine Meets Virtual Reality Conference(MMVR02), January 23-26, Newport Beach, CA, United States of America.

Willans, J.S., Smith, S.P., Harrison, M.D. (2001) *Using scenarios to identify the design requirements of virtual environments*, Technical Report YCS 333, University of York, United Kingdom.

Willans, J.S., Harrison, M.D. (2001) A toolset supported approach for designing and testing virtual environment techniques, *International Journal of Human-Computer Studies*, 55, 145-165

Wood, L.E. (1996) the Ethnographic interview in user-centred work/task analysis. Wixon, D., Ramey, J. (Eds.), *Field methods casebook for software and systems design*, John Wiley and Sons, New York, United States of America, 35-56.

Zeltzer, D. (1992) Autonomy, interaction and presence. *Presence*, 1(1), 127-132.

Zudilova, E.V., Sloot, P.M.A., Belleman R.G. (2002) A Multi-modal Interface for an Interactive Simulated Vascular Reconstruction System, *Proceedings Of The 4th IEEE Int. Conference On Multimodal Interfaces*, Los Alamitos, California, United States of America.

Zudilova E.V., Sloot P.M.A. (2003) Virtual Reality and Desktop as a Combined Interaction-Visualisation Medium for a Problem-Solving Environment, *Proceedings of the International Conference on Computational Science*, Melbourne, Australia.

Appendices

1: Interview questions (Dutch)
2: Example of interacting with the VRE
3: Patient trajectory and information flow in vascular disorders diagnosis and treatment planning

8. Exploratory Interview

Exploring the process of vascular disorder diagnosis and treatment from patient intake to patient monitoring. To be conducted with 4 subjects involved in the process of diagnosis and treatment of vascular disorders. Interviews are conducted in Dutch to aid both interviewer and interviewees. Interviews are taped.

9. Subject functie en taken

1. Wat is uw functieomschrijving?

2. Wat zijn de taken die u binnen deze functie vervult?

3. Wat zijn uw verantwoordelijkheden? (*aansprakelijkheid*)

4. Bij welke afdeling(en) bent u werkzaam?

10. Proces diagnose en behandeling vasculaire stoornissen

5. Kunt u een beschrijving geven van het hele traject dat een patiënt doorloopt bij het vermoeden van een vasculaire stoornis? Uit welke taken bestaat het traject?

6. Welke personen en afdelingen spelen een rol in het traject en wat zijn hun functies?

11. Vasculaire stoornissen

7. Welke categorieën van vasculaire stoornissen zijn er en welke behandelt u? Hoe wordt elke stoornis gediagnostiseerd, welke expertregels zijn er? Welke testen matchen met vermoedens van specifieke stoornissen? Welke metingen worden daarbij gedaan?

8. Welke (categorieën) behandelmethoden bestaan voor vasculaire stoornissen? Wanneer worden welke methoden toegepast?

12. Informatie

9. Welke informatie wordt in de taken van het besproken traject gebruikt, waar komt deze vandaan, waar gaat de informatie naartoe en in welk formaat wordt deze informatie nu gebruikt?(van wie, welk systeem, welk formaat)?

    - Welke visuele informatie gebruikt u?

    - Welke numerieke data gebruikt u?

10. Bij welke van de taken in het besproken traject worden scans gebruikt? wat voor scans (MRI?MRA?CT?), wat is precies gescand, hoe wordt de scan gevisualiseerd, welke informatie wordt hier uit gehaald, hoe wordt de informatie gepresenteerd bij welke gelegenheid aan wie?

13. Knelpunten

11. Waar liggen de risico's in het proces van diagnose en planning van behandeling van een patiënt met een vasculaire stoornis; Waar in het proces

worden fouten met de meeste frequentie gemaakt, en waar in het proces worden de ernstigste fouten gemaakt?

14. Mentale verwerking visualisatie

12. Hoe visualiseert u zelf mentaal de gebruikte visuele informatie?

13. Heeft u behoefte aan (nieuwe) 3D visualisatie? Bij welke taken en scans zou dit handig zijn?

15. Systemen

13. Welke systemen worden gebruikt voor diagnose van vasculaire stoornissen, of kent u?
    In hoeverre speelt 3D visualisatie een rol?
    Wat zijn de grootste pluspunten van deze systemen?
    Wat zijn de grootste nadelen van deze systemen?

14. Welke systemen worden gebruikt voor planning van behandelingen, of kent u?
    In hoeverre speelt 3D visualisatie een rol?
    Wat zijn de grootste pluspunten van deze systemen?
    Wat zijn de grootste nadelen van deze systemen?

15. Heeft u behoefte aan meer informatie over blood flow? Ziet u mogelijkheden voor gebruik van bloodflow simulatie?

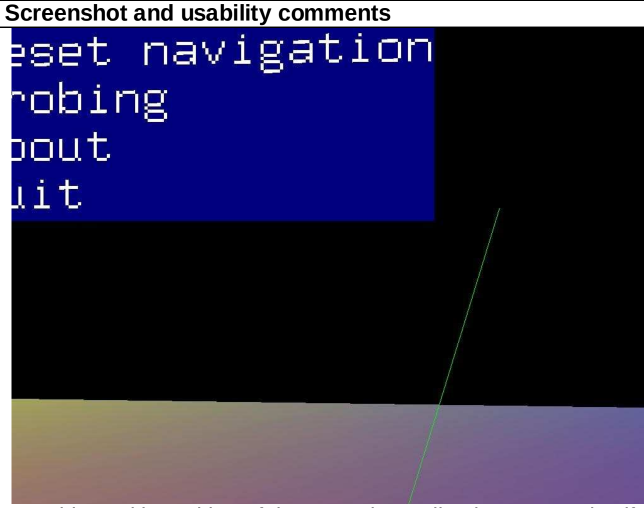
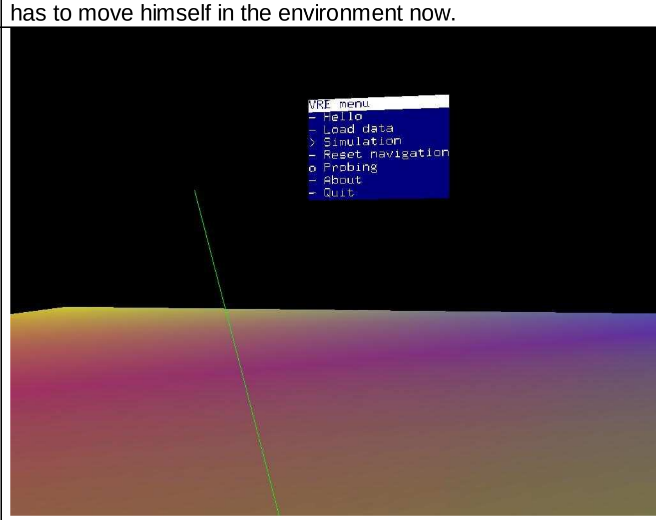
16. Example interaction with the VRE

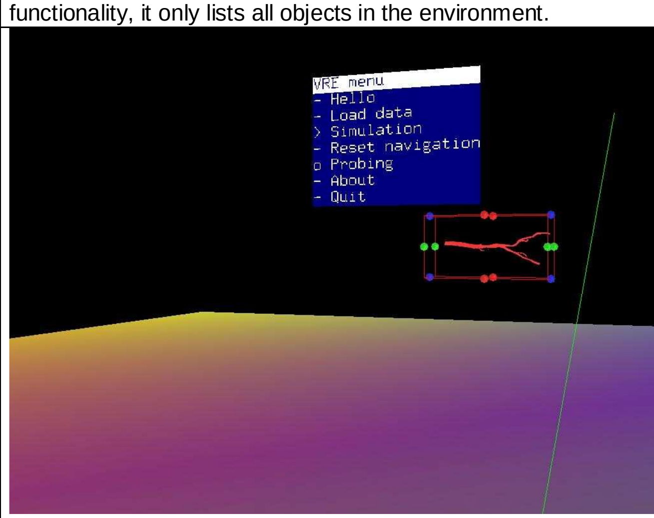
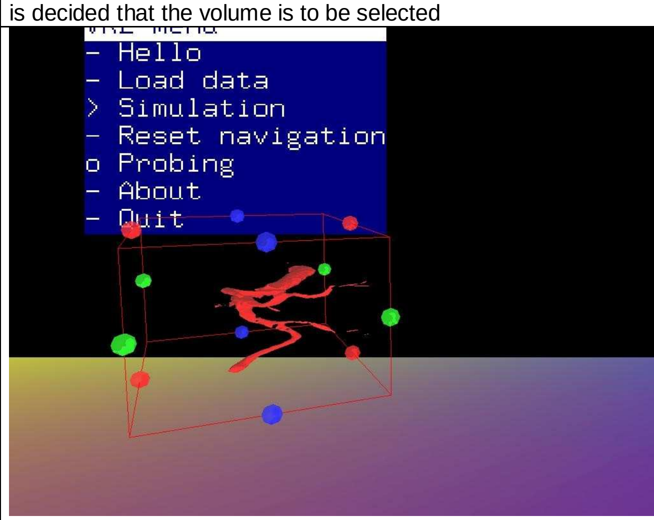
Task: Inspecting an artery

The goal of the user in this example is to inspect out- and inside of an artery. Data is loaded, the volume (VRE terminology for the loaded data object, in this case the aortic object) is shown and the user navigates the environment, sets the volume's size and explores in and outside of the object.

Problems are apparent mainly with the required amount of navigation of the environment before any inspection of the artery can actually take place. Even while the vessel is the object of interest, the user has often to move himself, instead of moving and rotating the volume to be able to inspect it.

Voice recognition problems and jitter of the tracking system add to the difficulty using the system.

user action	system reaction	Screenshot and usability comments
0. <startup procedure>	<visual>: Menu is shown in virtual environment along with pointer (green line in screenshot).	 A problem with position of the menu immediately presents itself to the user. The menu can be freely moved around in the environment, but unfortunately the top of the menu which has to be clicked and held to move the menu, is not yet visible. The user has to move himself in the environment now.
1. <Space mouse>: Push back, push left to move own position	<visual feedback>: User position changed	 The menu is now completely visible, its orientation is not yet ideal, but manageable.
2. <voice>:	A 'volume' (artery object) is	The volume is shown, off-centre in the screen,

'load data'	created	
3. <voice>: 'select volume'	<voice feedback>: 'object not found'	Incorrect voice command interpretation 'Help' is available, but instead of offering help on e.g. functionality, it only lists all objects in the environment.
4. Repeated: <voice> 'select volume'	<Voice feedback>: 'volume selected' <Visual feedback>: Selection is shown by bounding box with green, blue and red markers	 The term 'volume' needs to be familiar to users, before they can use the voice command; this does have to do with the unfamiliar term 'volume', but is also general problem for voice commands which have to be remembered rather than just recognized. Selection markers can be clicked and held to change object size and orientation, the functionality of these markers is however unclear to the user unfamiliar with the VRE. Alternative user action: <space mouse> point-and-click volume, is at this time however very difficult because of intense jitter in pointer which makes selection of a menu item at this distance of the menu very challenging.
5. <voice>: bigger	<voice feedback>: 'size 120%' <visual feedback>: size of the volume set to 120%, position of volume changed to the right and upper position of the screen	Setting size using discrete steps seems more difficult than an 'continuous' size change using dragging markers, however the <space mouse> point-and-click marker and dragging until the desired size of the object is reached, is very difficult because of the intense pointer jitter. The off-centre changing of position of the volume after setting its size is inconvenient and requires repositioning the volume. Here is decided that the volume is to be selected
6. <space mouse>: click and hold middle button while dragging the volume to move the volume <space mouse>: click a green marker and hold middle button while dragging the volume to rotate the volume	<visual feedback>: volume moved and re-oriented	 Inconsistent with moving the volume using voice commands, now no wireframe is shown around the volume while moving the volume. The volume has been selected using pointer in above

		screenshot in order to rotate the object. To explore the vessels further, the user has now got to move towards the object, rotate it further and move into the object to explore the inside.
7. <Space mouse>: Push back, push left to move own position inside the artery	User position changed	<<cap not available due to mediaplayer /print screen conflict, does the reader have trouble understanding this bit now?>> Moving inside and staying within the artery is difficult. Because there is no restriction on movement, the user can finds himself moving through the artery wall, making inspection of the limited space inside the vessel very difficult. This is made worse by the jitter in movement and various uncomfortable physical postures needed to hold the space mouse in such a way that the pointer stays within the vessel. Current resolution of the object is not high enough for a meaningful inspection of the artery.